

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2002, 12:10:12 ; Search time 57.07 Seconds  
(without alignments)  
903.071 Million cell updates/sec

Title: US-09-863-824-2

Perfect score: 2517

Sequence: 1 MVRLAAELLGLLLTLH.....CTSPSDEYIKQFQAREY 464

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_032802.\*
- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
  - 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
  - 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
  - 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
  - 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
  - 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
  - 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
  - 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
  - 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
  - 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
  - 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
  - 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
  - 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
  - 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
  - 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
  - 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
  - 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
  - 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
  - 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
  - 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
  - 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
  - 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1353	53.8	246	22 AAM25462	Human protein sequ
2	1035	41.1	198	21 AAB41922	Human ORFX ORF1686
3	927	36.8	571	22 AAB88393	Human membrane or
4	927	36.8	571	22 AAB49765	Human proliferatio
5	836	33.2	149	22 ABG06346	Novel human diagno
6	315	12.5	215	22 ABG06347	Novel human diagno
7	225.5	9.0	914	22 ABG27398	Novel human diagno
8	182	7.2	156	22 AAU21208	Human novel foetal
9	143	5.7	1572	20 AAW99301	Human BAI2 protein
10	137.5	5.5	1584	20 AAW99300	Human BAI1 protein
11	134.5	5.3	1522	20 AAW99302	Human BAI3 protein

12	134.5	5.3	1522	21 AAB23601	Human secreted pr9
13	133.5	5.3	763	22 ABB62815	Drosophila melanog
14	131.5	5.2	1934	22 AAB72301	Human ADAMTS-9 alt
15	130	5.2	1328	22 ABG22373	Novel human diagno
16	129	5.1	1206	19 AAW47030	Bovine N-proteinase
17	128.5	5.1	242	21 AAB00040	Human COMP/TSP-1 c
18	128	5.1	239	14 AAR40823	Human thrombospond
19	128	5.1	1211	19 AAW47028	Human N-proteinase
20	127	5.0	57	20 AAY49505	Human METH1 thombo
21	127	5.0	57	22 AAB50007	TSPI domain #3. H
22	127	5.0	555	22 AAU02914	Angiotensin conver
23	127	5.0	731	22 AAU02913	Angiotensin conver
24	127	5.0	1152	21 AAB00042	Human thrombospond
25	127	5.0	1170	22 AAB90800	Human shear stress
26	127	5.0	1170	22 AAB74450	Human variant thro
27	126	5.0	206	21 AAB08135	Amino acid sequenc
28	125.5	5.0	873	22 ABB66441	Drosophila melanog
29	125	5.0	4561	22 ABG30203	Novel human diagno
30	125	5.0	9222	22 ABG21064	Novel human diagno
31	124	4.9	969	21 AAY53900	Amino acid sequenc
32	124	4.9	1048	22 AAB85695	Larynx carcinoma a
33	124	4.9	1203	22 AAB50004	Bovine metalloprot
34	124	4.9	2150	21 AAY53898	Amino acid sequenc
35	124	4.9	2165	22 AAB90617	Human secreted pro
36	123	4.9	157	21 AAB08133	Amino acid sequenc
37	122	4.8	837	20 AAW75425	Human aggrecan deg
38	122	4.8	837	21 AAY94429	Human PRO1563 (UNQ
39	122	4.8	837	22 AAG78228	Human aggrecanase-
40	122	4.8	837	22 AAU29199	Human PRO polypept
41	122	4.8	837	22 AAB66178	Protein of the inv
42	122	4.8	840	21 AAB21256	Human metalloprote
43	121.5	4.8	1882	22 AAB72286	Human ADAMTS-9 am
44	120.5	4.8	459	22 AAU02916	Angiotensin conver
45	118.5	4.7	1088	22 AAG67244	Amino acid sequenc

ALIGNMENTS

RESULT 1

AA25462

ID AAM25462 standard; Protein; 246 AA.

AC AAM25462:

DT 16-OCT-2001 (first entry)

XX Human protein sequence SEQ ID NO:977.

DE Human; cancer; HIV infection; human immunodeficiency virus;  
KW Antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiac; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia;  
KW antiaggagant; haemostatic; vulnery; antitumor; osteopathic; eczema;  
KW dermatologic; antiallergic; antidiabetic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; neurotropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder.

XX Homo sapiens.

OS WO200153455-A2.

PN 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

XX

PR	23-DEC-1999;	99US-0471275.	XX	Human OREF1686 polypeptide sequence SEQ ID NO:3372.
PR	21-JAN-2000;	2000US-0488725.	XX	Human; open reading frame: OREFX; detection; cytotropic; hepatotropic;
PR	25-APR-2000;	2000US-0552317.	KW	vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
PA	(HYSE-) HYSEQ INC.		KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
PI	Tang YT, Liu C, Drmanac RT;		KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX	WPI; 2001-457603/49.		KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX	N-PSDB; AAH99403.		KW	antiviral; antibacterial; antifungal; immunosuppressive; antithyroid;
DR	Isolated human polynucleotides encoding polypeptides, useful for the		KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
PT	treatment and diagnosis of e.g. cancer, ulcers and HIV infection -		KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX	Claim 20; Page 206; 1217pp; English.		KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX	AAH99166 to AAH99904 encode the human proteins given in AAH25225 to		KW	cholesterol ester storage; systemic lupus erythematosus; infection;
CC	AAH25963. The proteins can have activities based on the tissues and		KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
CC	cells they are expressed in, such as: antiinflammatory; antirheumatic;		KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
CC	antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;		KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;
CC	central nervous system; virucide; anti-HIV; fungicide; antimutagen;		KW	thrombosis; contraceptive.
CC	cardiovascular; antianaemic; antiagregant; haemostatic; vulnerary;		XX	Homo sapiens.
CC	antitumor; osteopathic; dermatological; antiallergic; antisthmatic;		OS	WO200058473-A2.
CC	antiparkinsonian; and immunostimulant. The proteins and polynucleotides		PN	05-OCT-2000.
CC	encoding them can be used in gene therapy, antisense therapy and vaccine		PD	31-MAR-2000; 2000WO-US08621.
CC	production, the proteins and polynucleotides are useful for screening for		XX	31-MAR-1999; 99US-0127607.
CC	agonists or antagonists of a protein and for the treatment and diagnosis		PR	02-APR-1999; 99US-0127636.
CC	of disorders associated with the activity of a protein e.g. inflammation,		PR	05-APR-1999; 99US-0127728.
CC	rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,		PR	30-MAR-2000; 2000US-0540763.
CC	neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal		XX	(CURA-) CURAGEN CORP.
CC	infections, autoimmunity, genetic diseases, haematopoietic disorders,		PA	Shimkets RA, Leach M;
CC	anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,		XX	WPI: 2000-602362/57.
CC	osteoporosis, severe combined immunodeficiency, eczema, allergic		PI	N-PSDB; AAC76131.
CC	rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,		XX	Novel nucleic acids and peptides derived from open reading frame X,
CC	Alzheimer's disease, Parkinson's disease, neurodegenerative and		PT	useful for treating e.g. cancers, proliferative disorders,
CC	neurological disorders.		PT	neurodegenerative disorders and cardiovascular disease -
XX	Sequence 246 AA;		XX	Claim 11; Page 2553-2554; 5507pp; English.
SQ			XX	AAC74446 to AAC77606 encode the proteins given in AAC40237 to AAC43397,
	Query Match 53.8%; Score 1353; DB 22; Length 246;		CC	which represent the human OREFX open reading frames 1 to 3161. The OREFX
	Best Local Similarity 99.6%; Pred. No. 1.5e-120;		CC	sequences have activities such as: cytotropic; hepatotropic; vulnery;
	Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		CC	antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
QY	175 DQYXVDSNDDSNFLNPPRGNDHTAPGHRFTETKDQPEYDSTDGEGWLSVSGSVTCG 234		CC	osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
Db			CC	immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
	1 dQdykydstddsnfnpprgwdhtapghrtfctkdqpeydstgdsfwsvcsvtcg 60		CC	antidiabetic; hypotensive; dermatological; immunosuppressive;
QY	235 NGQKRTSCGYACTATESRCDRPNCPGIEDTFTTAATEVSLLAGSEEFNATKLFVDVT 294		CC	antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
Db			CC	antithyroid; and antianaemic. The sequences can be used for determining
	61 ngqkrtscgyactatesrctdrpncpgiedtfttaatevslagseefnatklfevdt 120		CC	the presence of or predisposition to, or preventing or treating
QY	295 DSCERWMSCKSEFLKYMVKMNDLPCSCSYTEVAYSTADIEDFKRKDFRWKDSGP 354		CC	pathological conditions associated with an OREFX-associated disorder. The
Db			CC	nucleic acids can be used to express OREFX proteins in gene therapy
	121 dscerwmsckseflkymhkvmdlpsccsypteavystadifdrkrdfwkdasgp 180		CC	vectors. The proteins and nucleic acids may be used to treat cancers,
QY	355 KEKLEIYKPARVCIRSMLSLESTTLAAOHCYGDNMQLITRGKAGTGNLSTEFSAEL 414		CC	proliferative disorders, neurodegenerative disorders, diabetes mellitus,
Db			CC	graft vs host disease, cardiovascular disease, cholesterol ester storage, systemic lupus
	181 kekleiykparvcirsmlesleatllaahqhcycgdmqlitrgkgagtgnllstefsael 240		CC	hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
QY	415 HYKVDV 420		CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
Db			CC	bacterial or fungal infection, malaria, autoimmune disorders, asthma,
	241 hykvdv 246		CC	allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
			CC	nocturnal haemoglobinuria, antiinflammatory disease; to enhance
			CC	coagulation; to inhibit thrombosis; and as a contraceptive.
RESULT	2		XX	Sequence 198 AA;
AAH41922			XX	Query Match 41.1%; Score 1035; DB 21; Length 198;
XX	AAH41922 standard; Protein: 198 AA.		XX	Best Local Similarity 97.3%; Pred. No. 2.4e-90;
AC	AAH41922;		XX	Matches 183; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
DT	08-FEB-2001 (first entry)			

Qy 60 SFSLSKEAPREHLHQAAHQPPRRFRQETGHPFSLQDRFSLFLDLPNFPDLSKADIN 119  
 Db 11 afalakeaprehldhqaahqpprrfrqetghpslqdrfslfldlpnfpdlskadin 70  
 Qy 120 QXNINQVIEVVGDPDSEADKQHPENKPSWSVSPDWRAMQSRSLARANSQDQYX 179  
 Db 71 gqnpnigvievvgdpdseadkqhpknkpswspspdrwagrsislaransgddqyk 130  
 Qy 180 YDSTSDSNFLNPPRGWDHTAPGHRTPETKDQPEYDSDGEGDWSLWSVCVSVTCGNGNQK 239  
 Db 131 ydstsdnflnpprgwdhtapghrtfctkdqpeydsdgedwslwsvcsvctcngngnk 190  
 Qy 240 RTRSCGYA 247  
 Db 191 rtrscgya 198

RESULT 3  
 AAB88393  
 XX AAB88393 standard; Protein; 571 AA.  
 AC AAB88393;  
 XX  
 DT 23-MAY-2001 (first entry)  
 XX  
 DE Human membrane or secretory protein clone PSEC0137.  
 XX  
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
 KW rheumatoid arthritis; diabetes.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI067182-A2.  
 XX  
 PD 10-JAN-2001.  
 XX  
 PF 07-JUL-2000; 2000EP-0114090.  
 XX  
 PR 08-JUL-1999; 99JP-0194179.  
 PR 11-JAN-2000; 2000JP-0118775.  
 PR 02-MAY-2000; 2000JP-0183766.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
 XX  
 DR WPI; 2001-093989/11.  
 DR N-PSDB; AAF93820.  
 XX  
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 PT gene therapy or as candidate target molecules in drug development -  
 XX  
 PS Claim 1; SEQ ID 154; 609pp + CD ROM; English.  
 XX  
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by  
 CC AAB88317 - AAB88419. Included in the invention are primers  
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the  
 CC cDNA sequences of the invention. The invention also includes methods for  
 CC the production of antibodies directed against the proteins, and cDNA  
 CC sequences, which can be used in vaccines. The polynucleotide sequences  
 CC can be used in gene therapy. The polynucleotide sequences and the  
 CC proteins they encode may be used in the prevention, treatment and  
 CC diagnosis of diseases associated with inappropriate secretory  
 CC protein/membrane protein expression. The nucleic acids and complementary  
 CC sequences may also be used as DNA probes in diagnostic assays  
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the  
 CC presence of similar nucleic acid sequences in samples. They may also be  
 CC used to study the expression and function of secretory proteins/membrane  
 CC polypeptides and their role in metabolism. The polypeptides may be used  
 CC as antigens in the production of antibodies against them and in assays to  
 CC identify modulators (agonists and antagonists) of expression and  
 CC activity. The antibodies and antagonists may also be used as therapeutic

CC agents to down regulate expression and activity. The antibodies may also  
 CC be used as diagnostic agents for detecting the presence of the  
 CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay  
 CC (ELISA). Examples of diseases which may be treated include rheumatoid  
 CC arthritis and diabetes.  
 XX  
 SQ Sequence 571 AA;  
 Query Match 36.8%; Score 927; DB 22; Length 571;  
 Best Local Similarity 42.7%; Pred. No. 2.3e-79;  
 Matches 200; Conservative 62; Mismatches 130; Indels 76; Gaps 17;  
 Qy 60 SFSLSKEAPREHLHQAAHQPPRRFRQETG-----HPSLQRDF----- 99  
 Db 117 stlistpnp-----dtqasaspdrprlreeearllprthlaelhqhcwtvtepaaltp 172  
 Qy 100 -----PRS-----FLDLNPNFDLSKADINQXPNQVIEVVGDPDSEADKQ--QHPPN 147  
 Db 173 gnatpprtqevtpilllelqkipeivhatlstpnqndqvtikvvedpqaevsidllaepsn 232  
 Qy 148 KP-----SWSVSPDWRAMW-----QRSLSLARANSQ-----DQDY-----XYDSTS 184  
 Db 233 pppqdtlsw-lpal-wpflwgdykgeekdrapgekeededypsedlegedqdekee 290  
 Qy 185 DDSN-----FLNPPRGWDH--TAPGHRTPETKDQPEYDSDGEGDWSLWSVCVSVTCGNGNQ 238  
 Db 291 deeeqalwfnqtdnwdqglapgdwvf--kdsvsyd-yepqkewspwpcsgncstqkq 347  
 Qy 239 KTRSCGYACTATESRTCDRNCPCGIE--DTFRTAATVSLLAGSEEFNATKLFVDVDS 296  
 Db 348 qtrpcgygctatetrtcdlpscpgetdkdtlgipseeekillar----natdmhdqdvds 403  
 Qy 297 CERWMSKSEFLKYMVKVMNDLPCSCPSYPTVAVYATADIFDRKRDPRKWDASGPKE 356  
 Db 404 cekwlncsksfliqysqmlrdlpscpaypleamdspsvsgdehqgfrwrdasgpre 463  
 Qy 357 KLEIYKPTARYCIRSMLESLESTTLAAOHCCYGDNNMOLITRGKGAGCTPNLISTEFAELHY 416  
 Db 464 rldlyqptarfclrmslsgesstlaahccydederslltrkgagmpnlstdfspklnhf 523  
 Qy 417 KVDVLPWIICKGWSRYNEARPPNNGOKCTESPSEDYIKQFQEAERY 464  
 Db 524 kfdttplwlcgdsrhlhavlpnngnrractdnpleeeylaqlqaekey 571

RESULT 4  
 AAB49765  
 ID AAB49765 standard; Protein; 571 AA.  
 AC AAB49765;  
 XX  
 DT 20-APR-2001 (first entry)  
 XX  
 DE Human proliferation differentiation factor amino acid sequence.  
 DE Human; proliferation differentiation factor; haematopoietic function.  
 KW Homo sapiens.  
 OS  
 PN WO200104312-A1.  
 XX  
 PD 18-JAN-2001.  
 XX  
 PF 06-JUL-2000; 2000WO-JP04514.  
 XX  
 PR 08-JUL-1999; 99JP-0194179.  
 PR 18-OCT-1999; 99US-0159586.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Yoshida K, Masuho Y;  
 XX

DR WPI: 2001-138354/14.  
DR N-PSDB; AAF29348.  
XX  
XX proliferation-differentiation factor protein encoded by PSEC137 cloned  
PT from human cDNA library, being hematopoietic factor for inducing  
PT differentiation of blood cells, used to maintain hematopoietic function  
PT -  
XX  
XX Claim 1; Page 42-44; 49pp; Japanese.  
PS  
XX This invention relates to polynucleotide sequence PSEC137, which encodes  
CC a human proliferation differentiation factor protein. Included in the  
CC invention is a vector containing the cDNA sequence, a transformant  
CC containing the vector, and a process for producing the protein. The  
CC protein together with its encoded DNA and antisense DNA are used in drugs  
CC to maintain the haematopoietic function. The present sequence represents  
CC the human proliferation differentiation factor protein amino acid  
CC sequence.  
XX  
XX Sequence 571 AA;  
SQ

Query Match 36.8%; Score 927; DB 22; Length 571;  
Best Local Similarity 42.7%; Pred. No. 2.3e-79;  
Matches 200; Conservative 62; Mismatches 130; Indels 76; Gaps 17;  
QY 60 SFSLSKEAPREHLHQAAHQPPRRFRQETG----HPSLORDF----- 99  
Db 117 stlstatpn-----dtqasapdrplreeeearllprthlqaelhqcwtvtepaalt 172  
QY 100 -----PR-----FLIDLFPNFDLSKADINGQXPNQIVTIEVDPDSEADKD--QHPEN 147  
Db 173 gnatprtqvtplllelqklpelvhatlstpnqnvtkvvedpqaevsidllaepsn 232  
QY 148 KP-----SNVSPDPRAWN-----ORSLSLARANS-----DQDY-----XYDSTS 184  
Db 233 pbbqdtlsw-lpal-wpflwgdykgeekdrapgekgeekedypsedlegedgedkee 290  
QY 185 DDSN-----FLNPPRGWDH--TAPGHRFTFKDQPEYDSTDGDSWLSVCSVTCGNQ 238  
Db 291 deeeqalwngttdnwgqwapgwff--kdsvsyd-yepqkewspwpcsgncstgkq 347  
QY 239 KTRTSCGYACTATSRCTDRPNCPIE--DTFRTAATEVSLAGSEEFNATKLFVEVDTS 296  
Db 348 qrtprcgygtatctdrlpscpgetdktlglpseewkllar-----natdmhdqgdvs 403  
QY 297 CERWMSCKSEFLKKYMHKVMNDLPSCPCSYPTVEAYSTADIFDRIKRDWRKDSGPKE 356  
Db 404 cekwlncskdflkylsqmldrlpspcpcaypleamdpsvlgdehqgrsfrwrdasgpre 463  
QY 357 KLEYKPTARYCIRSMLSLESTTLAAQHCCYGDNMOLITRGAGCTPNLISTEFSAEIHY 416  
Db 464 rldiyqptarfclrmslsgesstlaaqhccydedsrlltrgkagmpnlstldfsgklhf 523  
QY 417 KVDVLPWIIKGDWSRYNEARPNPNCCKTESPDEDIYKQFQEAR 464  
Db 524 kfdtptwllckgdwsrlhavlppnngtractdnplseeeylaqlqaekey 571

RESULT 5  
ABG06346  
ID ABG06346 standard; Protein; 149 AA.  
XX  
XX ABG06346;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #6337.  
DE  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
KW  
XX Homo sapiens.  
OS

XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
PF 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
PI WPI; 2001-639362/73.  
XX N-PSDB; AAS70533.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX Claim 20; SEQ ID No 36705; 103pp; English.  
PS The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC a food supplement. (II) and its binding partners are useful in medical  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological actions  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 149 AA;  
SQ  
Query Match 33.2%; Score 836; DB 22; Length 149;  
Best Local Similarity 99.3%; Pred. No. 1.5e-71;  
Matches 148; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 175 DDYXYDSTSDSFLNPPRGWDHTAGHRTFETKDPQPEYDSTDGDSWLSVCSVTCG 234  
Db 1 ddykydstsdnsflnpprgwdhtapghrtfctkddpeydstdgedgswlsvcsvtcg 60  
QY 235 NGNOKRTSCGYACTATSRCTDRPNCPIEDTFTTAATEVSLLAGSEEFNATKLFVEVD 294  
Db 61 ngngkrtscgyactatesrtcdprncpgiedtfttaatevslilagseefnatklfevdt 120  
QY 295 DSCERWMSCKSEFLKKYMHKVMNDLPSCP 323  
Db 121 dscerwmsckseflkkyhmhkmndlpscp 149  
RESULT 6  
ABG06347  
ID ABG06347 standard; Protein; 215 AA.  
XX  
XX ABG06347;  
XX 13-FEB-2002 (first entry)  
DT



```

XX DE Novel human diagnostic protein #6338.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI: 2001-639362/73.
XX DR N-PSDB: AAS70534.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID No 36706; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 215 AA;

Query Match 12.5%; Score 315; DB 22; Length 215;
Best Local Similarity 38.6%; Pred. No. 1.3e-21;
Matches 76; Conservative 24; Mismatches 49; Indels 48; Gaps 7;

Qy 213 EYSTDGEGDWSLWVCSVTGNGNQKTRSCGYACTATESRTCDRPNCPGI-----E 265
Dy 20 eenstdgedwslwvcsvtcgngnqktrscgyactatesrtcdrpncpevlkscqght 79

Qy 266 DTFRTAATEVSLLAGSEFNATKLFVEYDTSCERMWCKSEFLKKYHKV-----M 316
Dy 80 qsltealtfpfilasgyrkhlk1-avplkncqrvt-----khpphkicpqtptlrev 132

Qy 317 NDLESCPCSYTEVAYSTADIFDKRKRFRWKDASGPKKELEYKTP-TARYCIRSMLSL 375
Dy 133 ddvdkhlgtpkppkldqttad-----lqydrptakealksrngm 171

Qy 376 ESTTLA---AQHCYGD 389

```

```

Db 172 slaglsiqngqrcswge 188
RESULT 7
ABG27398
ID ABG27398 standard; Protein: 914 AA.
XX AC ABG27398;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #27389.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI: 2001-639362/73.
XX DR N-PSDB: AAS91585.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID No 57757; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 914 AA;

Query Match 9.0%; Score 225.5; DB 22; Length 914;
Best Local Similarity 87.9%; Pred. No. 3.7e-12;
Matches 51; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

Qy 1 MVRLAELLLGLLLLLTHTLVLRGSGAAGDPDAAAGNASQAQLNNLVGSDTSE 58

```

Fri Apr 26 10:29:58 2002

us-09-863-824-2.rag

Db 487 mvrtaealllllllllthitvrlsgaadgpdagaaagaaqqlq-glllgkptlse 543

RESULT 8

AAU21208  
ID AAU21208 standard; Protein; 156 AA.

XX AC AAU21208;

XX DT 18-DEC-2001 (first entry)

XX DE Human novel foetal antigen, SEQ ID NO 1452.

XX KW Human; foetal tissue antigen; antiinflammatory; neuroprotective;  
immunomodulator; cardiovascular; cytostatic; nephroprotective;  
cardiovascular; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; breast neoplasm; cancer;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; food additive.

XX OS Homo sapiens.

XX PN WO200155312-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01321.

XX PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220964.  
PR 26-JUL-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0228668.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-024617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.



Db	577	sheyrylylsirehlagqrlagegmsqvr-----slqellarrtyysgdllifsvdil- 631	
Qy	423	WIICKGWSRYNEARPPNNGKCTEPSPDEDIYKQFQ 459	
Db	632	-----invtdtkratyypsaddvqffq 655	
RESULT 10			
ID	AAW99300	AAW99300 standard; Protein; 1584 AA.	
XX	AC	AAW99300;	
XX	DT	21-MAY-1999 (first entry)	
XX	DE	Human BAI1 protein.	
XX	KW	Human; BAI1; brain; cancer; drug; diagnosis; prevention; treatment.	
XX	OS	Homo sapiens.	
XX	PN	JP11032766-A.	
XX	PD	09-FEB-1999.	
XX	PF	16-JUN-1997; 97JP-0176485.	
XX	PR	23-MAY-1997; 97JP-0150460.	
XX	PA	(SAKA ) OTSUKA PHARM CO LTD.	
XX	WPI	1999-183823/16.	
DR	N-PSDB	AAAX21355.	
XX	PT	New human BAI gene - is expressed in brain plays important role in cancer formation	
XX	PS	Claim 1; Page 18-22; 62pp; Japanese.	
XX	CC	This sequence represents the human BAI1 protein. The gene is expressed specifically in the brain and plays an important role in cancer formation in the brain. The BAI1 protein can be used in drug compositions to diagnose, prevent or treat such cancers.	
XX	SQ	Sequence 1584 AA;	
Query Match 5.5%; Score 137.5; DB 20; Length 1584;			
Best Local Similarity 21.8%; Pred. No. 0.0021;			
Matches 119; Conservative 50; Mismatches 177; Indels 201; Gaps 30;			
Qy	2	VRLAELLGLLGLLLTLHTVLRGSGADGPDAA-----GNASQAQLNNLNVGCS 53	
Db	11	vvilaplllll--llllgraraagadagpcepccatlvggkffgyfsaaa-----vfp 62	
Qy	54	DTTSETSFSLSKAPREH-LDHOAAHQPPRPFRFRETCHGHSQLRDFP-RSFL-----L 105	
Db	63	anasrcswtlrnpdrrtytlmkvakpvp-----csgpgrvrtvqfdfslestrtyl 115	
Qy	106	DLNFPD-LSKADINGQXPNIQVITIEVDPDSEADKDQHPEN---KPSWSVPSP--DW- 158	
Db	116	gvesfdevlrlcdpsaplaflqskqfl-----qmrqppqhdglrpragppgptddfs 170	
Qy	159	-----RA-----WQRSLSLARAN-----SGDQDYXYDSTSD 186	
Db	171	veylvgnrnpnsraacqmlcrwldaclagrsrshpcgimtpcaci9geag----- 221	
Qy	187	SNFLNP-----PRG-----WDHTAPGHR---TFETKDQPEYDSTGDGWSLWCVSVTC 233	
Db	222	-----gpaagplargdvclrdavaggpenceltsitdrgghgatggkwlswlgectrdc 277	
Qy	234	GNGNQKTRSC-----GYACTAT--ESRTCDRPNCPGIEDIFRTAATEVSLLA----- 279	

Db	278	ggglqtrtrtclpapvgveggcgcvleegrqcnreac---gpagrtssrsqslrstdarr 334	
Qy	280	-----GSE-----EFNATKLEVDTDSCERWMSCKSEFLKYYMHKVMNDLPSCCPCSYPTVA 331	
Db	335	reelgdqlqfgfpapqtgdpaeeewspwsvcss-----tcg----- 371	
Qy	332	YSTADIFDRIKRRKDFRKDASGPKKELEYKPTARYCIRSMLSLE-STTLAAQHCCYGDN 390	
Db	372	-----egqtrtrfcvssysystgcsgplreqrlc---n 401	
Qy	391	MLITRGKAGTPNLISTEFSAEHLHYKVDVLPWIIIC-----KGDWSRYNEARPPNNGKQC 445	
Db	402	nsavcpvhgaw-----dews-----pwslcstcgrgfrdrtrtcrppqfggnp 445	
Qy	446	TESPSDE 452	
Db	446	cegpekq 452	

RESULT 11

AAW99302

ID AAW99302 standard; Protein; 1522 AA.

XX AC AAW99302;

XX DT 21-MAY-1999 (first entry)

XX DE Human BAI3 protein.

XX KW Human; BAI3; brain; cancer; drug; diagnosis; prevention; treatment.

XX OS Homo sapiens.

XX PN JP11032766-A.

XX PD 09-FEB-1999.

XX PF 16-JUN-1997; 97JP-0176485.

XX PR 23-MAY-1997; 97JP-0150460.

XX PA (SAKA ) OTSUKA PHARM CO LTD.

XX WPI; 1999-183823/16.

DR N-PSDB; AAX21357.

XX PT New human BAI gene - is expressed in brain plays important role in cancer formation

XX PS Claim 2; Page 26-30; 62pp; Japanese.

XX CC This sequence represents the human BAI3 protein. The gene is expressed specifically in the brain and plays an important role in cancer formation in the brain. The BAI3 protein can be used in drug compositions to diagnose, prevent or treat such cancers.

XX SQ Sequence 1522 AA;

Query Match 5.3%; Score 134.5; DB 20; Length 1522;

Best Local Similarity 21.4%; Pred. No. 0.0038;

Matches 84; Conservative 48; Mismatches 170; Indels 91; Gaps 21;

Qy 62 SLUSKEA---PREHL---DHQAAHQPPRPFRFRETCHGHSQLRDFPFRSFLD-----LPN 109

Db 241 nltreakppkeefgmngdhtiksq---rprsvhekrypgeqadaak-fmaqtgesgvee 296

Qy 110 FFDLSKADIN-CQXPNIQVTIEV-----VDGPDSEADKDQHPENKPSWSVPSPDRAWW 162

Db 297 wsgwatcstvccgsgsvrtrtcvpygthcsqplresrvcmntalcphgv-----weews 352

Qy 163 QRSLSLARANSGDODYXYDSTSDSDSNFLNPPRGWDHTAPGHRFTFETKDQPEYDSTDGE-G 221

```
Db 353 pwslcstcgrgtrtrtrsc-----ppqyggprcpgpethhkhpcnialcpvqgqw 404
QY 222 DWSLWSVCSVTCGNGNQKTRSC-----GYACTA--TESRTCDRPNCPGIEDTFTTAAT 273
Db 405 ewsswqscvtsngtqgrsrqtaahgscrgpwaesrecynpec-----452
QY 274 EVSLLAGSEEFNATKLFVDVDSCE-----RWMSCKSEFLK-KYMHKVMNDLPSCP--C 324
Db 453 -----tangqwnqghwscskscdggwerrirtcgavltgqgcggtgeevrrcseqrc 507
QY 325 SYPTEVA---YSTADIFDRIKRDFRWK-----DASGPKK---LEIY-----KPTARYC 368
Db 508 papeicpedylnsmvwkrtpgdlafnqcpnatgttsrrcslnhgvafwqpsfarc 567
QY 369 IRS-----MLSLESTTLAAQHCCYGDNMQLITR 396
Db 568 lsneyrhlghsikehlaqgrmlagdgmsqvtk 600

RESULT 12
ID AAB23601 standard; Protein; 1522 AA.
AC AAB23601;
XX
DT 12-JAN-2001 (first entry)
DE Human secreted protein SEQ ID NO: 2.
KW Human; secreted protein; cytokine; cell proliferation;
KW nutritional supplement; immune modulation; autoimmune disorder;
KW haematopoiesis regulation; tissue growth; haemostasis; inflammation.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 3..15
FT Protein /label= signal_peptide
FT Protein 16..1522 /label= mature_protein
XX
PN WO200049134-A1.
XX
PD 24-AUG-2000.
XX
PF 18-FEB-2000; 2000WO-US04340.
XX
PR 19-FEB-1999; 99US-0120680.
PR 23-APR-1999; 99US-0298733.
PR 17-AUG-1999; 99US-0149639.
PR 23-SEP-1999; 99US-0155686.
PR 01-OCT-1999; 99US-0157247.
PR 29-NOV-1999; 99US-0167822.
PR 29-NOV-1999; 99US-0167823.
PR 15-FEB-2000; 2000US-0298733.
XX
PA (ALPH-) ALPHAGENE INC.
XX
PI Valenzuela D, Yuan O, Hoffman H, Hall J, Repiejko P;
XX
DR WPI; 2000-549267/50.
DR N-PSDB; AAA93101.
XX
PT New secreted proteins and polynucleotides encoding them, which are
PT derived from Homo sapiens, useful for therapy, diagnosis, and research,
PT as well as nutritional sources or supplements -
XX
PS Claim 9; Page 234-238; 309pp; English.
XX
CC The present sequence is the sequence of a human secreted protein. Its
CC cDNA was isolated from a foetal brain cDNA library. The proteins
CC and coding sequences of the invention can be used in the isolation of
```

```
CC similar genes and proteins, in the elucidation of their function in vivo,
CC and to treat a number of conditions. It is possible that they may have
CC uses as nutritional supplements, as cytokine or cell proliferation
CC factors, in immune modulation, where they may be used to treat immune and
CC autoimmune diseases, as haematopoiesis regulators (treating myeloid or
CC lymphoid cell deficiencies), in the promotion of tissue growth, they may
CC have chemokine or chemotactic activity, haemostatic or thrombolytic
CC activity, or anti-inflammatory activity.
XX
SQ Sequence 1522 AA;

Query Match 5.3%; Score 134.5; DB 21; Length 1522;
Best Local Similarity 21.4%; Pred. No. 0.0038;
Matches 84; Conservative 48; Mismatches 170; Indels 91; Gaps 21;

QY 62 SLSKEA---PREHL-----DHQAAHQFPFRFRQETGHSLSQRDFFRSFLLD-----LPN 109
Db 241 nltreakrppkesfgmmgdhtiksq---tprsvhekrvpqeqadaak-fmaqgtgesgvee 296
QY 110 FPDLSKADIN-GQXPNIQVETIEV-----VDGPDSEADKDOHPENKPSWSVSPDWRWW 162
Db 297 wsgwstcscvctcgqsgvtrtcvpygthcsqplresvncntalcpvhgv-----weews 352
QY 163 QRSLSLARANSQDDQXYDSTDDSNFLNPPRGWDHTAPGHRTFTETKDQPEYDSTDE-G 221
Db 353 pwslcstcgrgtrtrtrsc-----ppqyggprcpgpethhkhpcnialcpvqgqw 404
QY 222 DWSLWSVCSVTCGNGNQKTRSC-----GYACTA--TESRTCDRPNCPGIEDTFTTAAT 273
Db 405 ewsswqscvtsngtqgrsrqtaahgscrgpwaesrecynpec-----452
QY 274 EVSLLAGSEEFNATKLFVDVDSCE-----RWMSCKSEFLK-KYMHKVMNDLPSCP--C 324
Db 453 -----tangqwnqghwscskscdggwerrirtcgavltgqgcggtgeevrrcseqrc 507
QY 325 SYPTEVA---YSTADIFDRIKRDFRWK-----DASGPKK---LEIY-----KPTARYC 368
Db 508 papeicpedylnsmvwkrtpgdlafnqcpnatgttsrrcslnhgvafwqpsfarc 567
QY 369 IRS-----MLSLESTTLAAQHCCYGDNMQLITR 396
Db 568 lsneyrhlghsikehlaqgrmlagdgmsqvtk 600

RESULT 13
ABB62815
ID ABB62815 standard; Protein; 763 AA.
XX
AC ABB62815;
XX
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 15237.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
```



Db 1581 cvddn-----knevgardcvakrpyvdresclqpcyvwitge----- 1619  
QY 345 DFRWKDAS-----GPKKEL-----EIV---KPTARYCIRSMLSLESTTLAAQHCCY 387  
Db 1620 ---wsecvtcgkykqrlvscseiygkenyeysttincpgtqppsvhpcy 1670

RESULT 15  
ABG22373  
ID ABG22373 standard; Protein; 1328 AA.  
AC ABG22373;  
XX  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #22364.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS86560.  
DR  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX  
PS Claim 20; SEQ ID NO 52732; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1328 AA;

Query Match 5.28; Score 130; DB 22; Length 1328;  
Best Local Similarity 21.84; Pred. No. 0.0083;  
Matches 63; Conservative 26; Mismatches 102; Indels 98; Gaps 13;

QY 192 PPRGWDHTAFCHRTFETKD---QPEYDSTDGE-GDWLSWVSVCTCGNGNKRTRSC--- 244  
Db 543 ppldgtmcapgkchcfkgchciwltpdlikrdgswgawspfgscrtcgvgkfrtrqcdnp 602  
QY 245 -----GYACT--ATESRTCDRPNCPGIEDTFRTAATEVSLLAGSEEFNATKLFVEVDTDSC 297  
Db 603 hpanggrtcsglaydfqlcsrqcdpsladfr-----eeqc 638  
QY 298 ERWMSCKSEFLKYMVKVMDLPSCPCSYPTVEAYSTADIFDIKRKDFRW-----KDAS 352  
Db 639 rqw-----diyfehgdaqhhwlphehrda- 662  
QY 353 GPKEKLEIYKPTARYCIRSMLSLESTTLAAQHCCYGDNMQLITRG--KGAGTPNLISTEF 410  
Db 663 --kerchlyceste--tgevsvmkrmvhdgtcrsykdafslcvrgdcrkvgcdgvg--- 715  
QY 411 SAEHLHYKVDVLPWLTICKGDWSRYNEARPPNNGQKCTESPSDEDIYIKOFQ 459  
Db 716 sskqedkcgv-----cggdnshckvkv-----gtftrspkhhgylkmfe 754

Search completed: April 25, 2002, 14:46:27  
Job time: 9375 sec



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 10:12:56 ; Search time 53.57 Seconds  
(without alignments)  
6396.467 Million cell updates/sec

Title: US-09-863-824-1  
Perfect score: 1395  
Sequence: 1 atgtgtgccttgcggccga.....aagaggccagggaattattaa 1395

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCRUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50.6	3.6	1155	2	US-08-387-942C-20
2	50.6	3.6	12588	2	US-08-387-942C-1
3	49.4	3.5	1155	2	US-08-387-942C-7
4	49	3.5	1149	4	US-09-422-487-6
5	48.8	3.5	1155	4	US-09-422-487-4
6	48.8	3.5	1158	3	US-08-844-188-45
7	48	3.4	2220	4	US-08-765-907A-14
8	48	3.4	2888	4	US-08-765-907A-1
9	47.8	3.4	1155	2	US-08-387-942C-18
10	46.8	3.4	1051	4	US-09-245-041-10
11	46.8	3.4	2419	4	US-09-245-041-8
12	46.8	3.4	2625	4	US-09-245-041-18
13	46.8	3.4	4072	4	US-09-245-041-16
14	46.8	3.4	8589	4	US-09-245-041-14
15	46.8	3.4	8827	4	US-09-245-041-1
16	45.8	3.3	2713	3	US-08-804-439A-13
17	45.8	3.3	2713	3	US-08-720-229-13
18	44.6	3.2	1176	2	US-08-387-942C-17
19	44.6	3.2	1635	3	US-09-234-332-4
20	43.6	3.1	477	4	US-09-135-994-1
21	43.4	3.1	3624	1	US-07-951-715A-6
22	43.4	3.1	3624	2	US-08-459-448A-6
23	43.4	3.1	3624	3	US-08-459-595A-6
24	43.4	3.1	3624	3	US-08-459-504B-6
25	43.4	3.1	3624	3	US-08-459-444-6
26	43.4	3.1	3624	3	US-09-053-549-7
27	43.4	3.1	3624	4	US-09-547-422-6

28	42.8	3.1	1465	4	US-09-338-671-1	Sequence 1, Appli
29	42.6	3.1	3472	6	5244792-2	Patent No. 5244792
30	42	3.0	4377	3	US-08-911-853-28	Sequence 28, Appl
31	42	3.0	4377	4	US-09-479-409-28	Sequence 28, Appl
32	42	3.0	4377	4	US-09-479-453-28	Sequence 28, Appl
33	41.8	3.0	8854	3	US-09-053-549-1	Sequence 1, Appli
34	41.6	3.0	4162	2	US-08-459-448A-26	Sequence 26, Appl
35	41.6	3.0	4162	3	US-08-459-595A-26	Sequence 26, Appl
36	41.6	3.0	4162	3	US-08-459-504B-26	Sequence 26, Appl
37	41.6	3.0	4162	3	US-08-459-444-26	Sequence 26, Appl
38	41.6	3.0	4162	4	US-09-547-422-26	Sequence 26, Appl
39	41.6	3.0	4165	1	US-07-951-715A-26	Sequence 26, Appl
40	41.4	3.0	1155	2	US-08-387-942C-21	Sequence 21, Appl
41	40.8	2.9	844	4	US-08-998-416-349	Sequence 349, App
42	40.8	2.9	2004	1	US-08-471-033-18	Sequence 18, Appl
43	40.8	2.9	2004	2	US-08-471-044-18	Sequence 18, Appl
44	40.8	2.9	2004	2	US-08-463-483A-18	Sequence 18, Appl
45	40.8	2.9	2004	2	US-08-471-046A-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1  
US-08-387-942C-20  
; Sequence 20, Application US/08387942C  
; Patent No. 5939289  
; GENERAL INFORMATION:  
; APPLICANT: ERTESVAG, HELGA  
; APPLICANT: VALLA, SVEIN  
; APPLICANT: SKJAK-BRAEK, GUDMUND  
; APPLICANT: LARSEN, BJORN  
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES  
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP  
; STREET: P.O. BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22042  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/387,942C  
; FILING DATE: 09-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURPHY JR, GERALD M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1809-106P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-205-8000  
; TELEFAX: 703-205-8050  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-387-942C-20

Query Match 3.6% Score 50.6; DB 2; Length 1155;  
Best Local Similarity 46.1%; Pred. No. 0.009;  
Matches 170; Conservative 0; Mismatches 199; Indels 0; Gaps 0;  
Qy 896 gctggatgagctgcaaaagcgagttcttaagaagatcatgcacaaagtgatgaatgacc 955

Db	545	CGCGGGTGTTCGAGAACAGCTCTCGTACAAACACCGCCACGCTTCAACATCGTCA	604
Qy	956	tgccagctgcccctgctctacccactgagtgccctacagcagcgccgacatcttcg	1015
Db	605	CCAGCACCACGACTTGTCTGAGCAACAACTGCGCTACGGCAACGGCGCGCGGCC	664
Qy	1016	accgcataagcgcaagacttcctgctggaagagacccagcgcccaaggaagactgg	1075
Db	665	TGTTGGTGCAGCGCGGCTCTCCGACGTGCGGCACCCCTACGACATCCTGATCGACGGCG	724
Qy	1076	agatctacaagccactgcccgttaactgcatccgtctccatgctgctcctgagagaccca	1135
Db	725	CGCGCTACTACGACAAAGCGCTGGAAGCGTGCAGATCAAGATGGCCACGACGTCACCC	784
Qy	1136	cgctggcgccagcactgtgtctacggcgacacaacatgcagctccatcaccaggggcaagg	1195
Db	785	TGCAGAACGGCGAGATCTACGGCAACGGCTATACGGGGTGCAGTCTACGGCGCGGAGG	844
Qy	1196	ggcgggcgcccaactcactcagaccgagttctccgagactccactacaaggtgg	1255
Db	845	ATGTGCAGATCCTCGACAATACATACACGACAATTCGAGAAAGGTTCTCTACCGGAAA	904
Qy	1256	acgtctctgc 1264	
Db	905	TCCTCTGC 913	
RESULT 2			
US-08-387-942C-1			
; Sequence 1, Application US/08387942C			
; Patent No. 5939289			
; GENERAL INFORMATION:			
; APPLICANT: ERTESVAG, HELGA			
; APPLICANT: VALLA, SVEIN			
; APPLICANT: SKJAK-BRAEK, GUDMUND			
; APPLICANT: LARSEN, BJORN			
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES			
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE			
; NUMBER OF SEQUENCES: 52			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP			
; STREET: P.O. BOX 747			
; CITY: FALLS CHURCH			
; STATE: VA			
; COUNTRY: USA			
; ZIP: 22042			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/387,942C			
; FILING DATE: 09-MAY-1995			
; CLASSIFICATION: 435			
; ATTORNEY/AGENT INFORMATION:			
; NAME: MURPHY JR, GERALD M.			
; REGISTRATION NUMBER: 28,977			
; REFERENCE/DOCKET NUMBER: 1809-106P			
; TELEPHONE: 703-205-8000			
; TELEFAX: 703-205-8050			
; INFORMATION FOR SEQ ID NO: 1:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 12588 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; MOLECULE TYPE: DNA (genomic)			
; ORIGINAL SOURCE:			
; ORGANISM: Azotobacter vinelandii			
; STRAIN: E			

;	FEATURE:		
;	NAME/KEY:	CDS	
;	LOCATION:	290..1951	
;	FEATURE:		
;	NAME/KEY:	CDS	
;	LOCATION:	2227..6438	
;	FEATURE:		
;	NAME/KEY:	CDS	
;	LOCATION:	6702..9695	
;	FEATURE:		
;	NAME/KEY:	CDS	
;	LOCATION:	9973..12588	
;	US-08-387-942C-1		
Query Match			
Best Local Similarity 3.6%; Score 50.6; DB 2; Length 12588;			
Matches 170; Conservative 0; Mismatches 199; Indels 0; Gaps 0;			
Qy	896	gctggatgagctgcaaaagcgagttctttaagaagtacatgcacaaggtgatgaatgacc	955
Db	7246	CGGGGGTGTTCGAGAACACGCTCTCGTACAAACAGCAGCCGACGCTTCAACATCGTCA	7305
Qy	956	tgccagctgcccctgctctacccactgagtgccctacagcagcgccgacatcttcg	1015
Db	7306	CCAGCACCACGACTTGTCTGAGCAACAACTGCGCTACGGCAACGGCGCGCGGCC	7365
Qy	1016	accgcataagcgcaagacttcctgctggaagagacccagcgcccaaggaagactgg	1075
Db	7366	TGTTGGTGCAGCGCGGCTCTCCGACGTGCGGCACCCCTACGACATCCTGATCGACGGCG	7425
Qy	1076	agatctacaagccactgcccgttaactgcatccgtctccatgctgctcctgagagaccca	1135
Db	7426	CGCGCTACTACGACAAAGCGCTTGAAGCGTGCAGATCAAGATGGCCACGACGTCACCC	7485
Qy	1136	cgctggcgccagcactgtgtctacggcgacacaacatgcagctccatcaccaggggcaagg	1195
Db	7486	TGAGAACGGCGAGATCTACGGCAACGGCTATACGGGGTGCAGTCTACGGCGCGGAGG	7545
Qy	1196	ggcgggcgcccaactcactcagaccgagttctccgagactccactacaaggtgg	1255
Db	7546	ATGTGCAGATCCTCGACAATACATACACGACAATTCGAGAAAGGTTCTCTACCGGAAA	7605
Qy	1256	acgtctctgc 1264	
Db	7606	TCCTCTGC 7614	
RESULT 3			
US-08-387-942C-7			
; Sequence 7, Application US/08387942C			
; Patent No. 5939289			
; GENERAL INFORMATION:			
; APPLICANT: ERTESVAG, HELGA			
; APPLICANT: VALLA, SVEIN			
; APPLICANT: SKJAK-BRAEK, GUDMUND			
; APPLICANT: LARSEN, BJORN			
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES			
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE			
; NUMBER OF SEQUENCES: 52			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP			
; STREET: P.O. BOX 747			
; CITY: FALLS CHURCH			
; STATE: VA			
; COUNTRY: USA			
; ZIP: 22042			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/387,942C			
; FILING DATE: 09-MAY-1995			
; CLASSIFICATION: 435			
; ATTORNEY/AGENT INFORMATION:			
; NAME: MURPHY JR, GERALD M.			
; REGISTRATION NUMBER: 28,977			
; REFERENCE/DOCKET NUMBER: 1809-106P			
; TELEPHONE: 703-205-8000			
; TELEFAX: 703-205-8050			
; INFORMATION FOR SEQ ID NO: 1:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 12588 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; MOLECULE TYPE: DNA (genomic)			
; ORIGINAL SOURCE:			
; ORGANISM: Azotobacter vinelandii			
; STRAIN: E			



Db 530 gctgagaccctgctgcgtcatcatggtgaacgactccaagatcgacaagaacacccaga 589  
Qy 947 tgaatgacctgcccagctgcccctgctctctacccccactgaggtggcctacagcagcgccg 1006  
Db 590 tcaagaccacccctactatctctcaagaataacaaagtactggaacctgcgccaaggct 649  
Qy 1007 acatctcgaccgcatacaagcgaagacttccgctggaagcgcagcgccgccaagg 1066  
Db 650 ccaacgctccctctcccgaccagaaagcgcagctacgactcagagtgggcaccgaga 709  
Qy 1067 agaagctggagatataagcccactgcccgtactgctatccgctccatgctgctccctgg 1126  
Db 710 agaaccagaagaccacactatcaacacggctggcctgcagatcaacatgactcgggga 769  
Qy 1127 agagcaccacgctggcgacagcactgctgctacggcgacaaacatcagctcatcacca 1186  
Db 770 tgaagttcagtgccgagtggtggcgccgaccggagacatacaagcaccagctccaccg 829  
Qy 1187 ggggcaaggggcgccgacgcccacacctcatcagcaccgagttctccgcgagagctccact 1246  
Db 830 agagctgaagtgagtgactccaccggagaccagatcatgaccaagtaccaggagcact 889  
Qy 1247 acaagtggaagctcctgccc 1266  
Db 890 ccgagatcgacaacccgacc 909

RESULT 6  
US-08-844-188-45  
; Sequence 45, Application US/08844188  
; Patent No. 6127180  
; GENERAL INFORMATION:  
; APPLICANT: Narva, Kenneth E.  
; APPLICANT: Schnepf, H. Ernest  
; APPLICANT: Knuth, Mark  
; APPLICANT: Pollard, Michael R.  
; APPLICANT: Cardineau, Guy  
; APPLICANT: Schwab, George E.  
; TITLE OF INVENTION: Pesticidal Toxins  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606-6669  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/844,188  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/633,993  
; FILING DATE: 19-APR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sanders, Jay M.  
; REGISTRATION NUMBER: 39,355  
; REFERENCE/DOCKET NUMBER: MA-703C1  
; TELEPHONE: 352-375-8100  
; TELEFAX: 352-372-5800  
; INFORMATION FOR SEQ ID NO: 45:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1158 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)  
US-08-844-188-45  
Query Match 3.5%; Score 48.8; DB 3; Length 1158;  
Best Local Similarity 45.5%; Pred. No. 0.023;  
Matches 173; Conservative 0; Mismatches 207; Indels 0; Gaps 0;  
Qy 887 gctgtgagcgcgtgagtgagtgcaaaagcgcgagttcttaagaagtacatgcacaaggtga 946  
Db 530 GCTGGACCCCTCGTGGCGTGCATCATGTTCAAGACTTCCAAAGATCGACAAGAACACCCAGA 589  
Qy 947 tgaatgacctgcccagctgcccctgctctacccccactgaggtggcctacagcagcgccg 1006  
Db 590 TCAAGACCACCCCGTACTATCTTCAAGAAATACAAAGTACTGGAAACCTTCGCCAAGGGCT 649  
Qy 1007 acatctcgaccgcatacaagcgaagcacttccgctggaagcgcagcgccgccaagg 1066  
Db 650 CCAACGTGTCCTCTCCCGCACCAAGCGCAGCTACGACTACGAGTGGGGCACCAGAGA 709  
Qy 1067 agaagctggagatatacaagcccactgcccgtactgctatccgctccatgctgctccctgg 1126  
Db 710 AGAACCAAGAACACCATCATCAACACCGTGGGGCCTGCAGATCAACATCGACTCGGGGA 769  
Qy 1127 agagcaccacgctggcgacagcactgctgctacggcgacaaacatgagctcatcacca 1186  
Db 770 TGAAGTTCGAGGTGCCGGAGGTGGCGGCGCACCGAGGACATCAAGACCCAGCTCACCG 829  
Qy 1187 ggggcaaggggcgccgacgcccacacacacacacacacacacacacacacacacacacac 1246  
Db 830 AGGAGCTGAAGTGGAGTACTCCACCGAGACCAAGATCATGACCAAGTACCAGGAGCACT 889  
Qy 1247 acaagtggaagctcctgccc 1266  
Db 890 CCGAGATCGACAACCCGACC 909

RESULT 7  
US-08-765-907A-14  
; Sequence 14, Application US/08765907A  
; Patent No. 6352839  
; GENERAL INFORMATION:  
; APPLICANT: THIBAUT, Denis  
; APPLICANT: BLANC, Veronique  
; APPLICANT: BAWAS-JACQUES, Nathalie  
; APPLICANT: BLANCHE, Francis  
; APPLICANT: COUZET, Joel  
; APPLICANT: BARRIERE, Jean-Claude  
; APPLICANT: DEBUSSCHE, Laurent  
; APPLICANT: FAMECHON, Alain  
; APPLICANT: PARIS, Jean-Marc  
; APPLICANT: DUTRUC-ROSSET, Gilles  
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By  
; TITLE OF INVENTION: Mutasynthesis  
; FILE REFERENCE: Streptogramin genes  
; CURRENT APPLICATION NUMBER: US/08/765,907A  
; CURRENT FILING DATE: 1997-03-20  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 2220  
; TYPE: DNA  
; ORGANISM: Streptomyces pristinaespiralis  
US-08-765-907A-14  
Query Match 3.4%; Score 48; DB 4; Length 2220;  
Best Local Similarity 46.9%; Pred. No. 0.042;  
Matches 150; Conservative 0; Mismatches 170; Indels 0; Gaps 0;  
Qy 950 atgacctgccagctgcccctgctctctacccccactgaggtggcctacagcagcgccgaca 1009  
Db 1658 aggacgcccgcgtcaaggcctcctcgcgcgcccgaagacgcagcagagaacctga 1717











GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2002, 12:13:27 ; Search time 28.09 Seconds  
(without alignments)  
403.470 Million cell updates/sec

Title: US-09-863-824-2  
Perfect score: 2517  
Sequence: 1 MVRLAELLGLLLTLH.....CTSPSDEYIKQFQAREY 464

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/6CTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128	5.1	239	5	PCT-US93-01652-1
2	127	5.0	1170	1	US-08-313-288B-20
3	117.5	4.7	807	1	US-07-862-021B-10
4	117.5	4.7	807	1	US-08-313-288B-10
5	117.5	4.7	807	5	PCT-US93-03164-10
6	115.5	4.6	1156	3	US-08-996-083-1
7	115.5	4.6	1156	4	US-09-429-516-1
8	115.5	4.6	1156	4	US-09-429-516-3
9	115	4.6	788	2	US-08-918-914-4
10	115	4.6	1172	1	US-08-313-288B-19
11	113	4.5	218	3	US-08-985-526-1
12	113	4.5	441	3	US-08-985-526-3
13	111	4.4	469	1	US-08-313-288B-15
14	110.5	4.4	1785	4	US-09-341-587-3
15	106.5	4.2	568	1	US-07-862-021B-14
16	106.5	4.2	568	5	PCT-US93-03164-14
17	106	4.2	898	2	US-08-808-982-5
18	106	4.2	898	4	US-09-306-902A-5
19	105	4.2	802	1	US-07-862-021B-12
20	105	4.2	802	1	US-08-313-288B-12
21	105	4.2	802	5	PCT-US93-03164-12
22	104.5	4.2	559	1	US-08-313-288B-14
23	103	4.1	943	2	US-08-808-982-7
24	103	4.1	943	4	US-09-306-902A-7
25	101	4.0	60	1	US-07-646-531D-12
26	101	4.0	60	2	US-08-488-273-12
27	101	4.0	60	6	5426100-12

28	95.5	3.8	437	2	US-08-935-450-8	Sequence 8, Appl
29	93.5	3.7	392	2	US-08-799-173A-7	Sequence 7, Appl
30	93	3.7	1036	4	US-08-891-640-3	Sequence 3, Appl
31	92	3.7	23	1	US-07-646-531D-7	Sequence 7, Appl
32	92	3.7	23	2	US-08-488-273-7	Sequence 7, Appl
33	92	3.7	23	4	US-09-197-770B-13	Sequence 13, Appl
34	92	3.7	23	6	5426100-7	Patent No. 5426100
35	91.5	3.6	598	1	US-08-487-890A-97	Sequence 97, Appl
36	91.5	3.6	598	2	US-08-478-435-97	Sequence 97, Appl
37	91.5	3.6	598	2	US-08-337-483-97	Sequence 97, Appl
38	91.5	3.6	598	2	US-08-478-373-97	Sequence 97, Appl
39	91.5	3.6	598	3	US-08-474-671-97	Sequence 97, Appl
40	91.5	3.6	598	3	US-08-483-577A-97	Sequence 97, Appl
41	91.5	3.6	598	4	US-08-897-438-97	Sequence 97, Appl
42	91.5	3.6	598	4	US-08-637-654-97	Sequence 97, Appl
43	91	3.6	366	4	US-08-857-076-103	Sequence 103, App
44	91	3.6	486	3	US-08-746-559A-5	Sequence 5, Appl
45	91	3.6	516	3	US-08-746-559A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
PCT-US93-01652-1  
; Sequence 1, Application PC/TUS9301652  
; GENERAL INFORMATION:  
; APPLICANT: Bouck, Noel P.  
; APPLICANT: Polverini, Peter J.  
; APPLICANT: Good, Deborah J.  
; TITLE OF INVENTION: Method and Composition for  
; TITLE OF INVENTION: Inhibiting Angiogenesis  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut  
; STREET: 100 South Wacker Drive, Suite 960  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606-4002  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/01652  
; FILING DATE: 19930222  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/841.656  
; FILING DATE: 24-FEB-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/464.369  
; FILING DATE: 12-JAN-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pentress, Susan B.  
; REGISTRATION NUMBER: 31,327  
; REFERENCE/DOCKET NUMBER: 92005-PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312)-456-8000  
; TELEFAX: (312)-456-7776  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 239 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
PCT-US93-01652-1



```

;
; TYPE: AMINO ACID
;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-03164-10

```

## RESULT

QY 341 IKRDFWKDASGPKLEIYKPTARYCIRS 371  
: : : | | : : | | : | |  
Db 694 TIOMEPOFGAPCP-ETVQRKKCARCKLRS 723

RESULT 6  
US-08-996-083-1

```

US-08-996-083-I
Query Match 4.6%; Score 115.5; DB 3; Length 1156;
Best Local Similarity 25.8%; Pred. No. 0.01;
Matches 41; Conservative 13; Mismatches 50; Indels 55; Gaps 9;

Qy 154 PSP-----DWRWQ-----RSLSLARANS-----DDYXY 180
||| :||| :||| :||| :
Db 46 PSPALEDEEASEWTSFNVDHPGGDGFESLAIRFYGPARYCPRPLAERTTDWAL 105
||| :||| :||| :||| :

Qy 181 DSTSDSNFLNPPRG-W--DHTAFG-----HRTFEKQDPEYDSTDGEGDWSLWSVCS 230
||| :||| :||| :||| :

Db 106 PSAGVERVHLNPTRGFWCLNREQPRGRCSNYHVRFCPLEASW-----GAWGFWGPCS 159
||| :||| :||| :||| :

Qy 231 VTCGNGNOKRTRSC-----GYACTA--TESRTCDRPNCPG 263
||| :||| :||| :||| :

Db 160 GSCGPGRRLLRRHCPSDAGDACPGRLPEAOKCVRPCPG 198
||| :||| :||| :||| :

```

```

160  GSCGPGKRRKHRCFSPAGUACPKRPLEAKGCKVRKPCPG 139

```

RESULT 7  
 US-09-429-516-1  
 ; Sequence 1, Application US/09429516  
 ; Patent No. 6251389  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Magna, Holly  
 ; APPLICANT: Schaffer, Paul  
 ; APPLICANT: Lawton, Michael  
 ; APPLICANT: Yocum, Sue  
 ; APPLICANT: Mitchell, Peter  
 ; APPLICANT: Hutchinson, Nancy  
 ; APPLICANT: Murry, Lynn E.  
 ; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOS  
 ; TITLE OF INVENTION: PHOYDROLASE-2  
 ; NUMBER OF SEQUENCES: 3  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSeq for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/429,516

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/996,083
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0420 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1156 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SATPF1002
; CLONE: 422069
; US-09-429-516-3
```

```
Query Match 4.6%; Score 115.5; DB 4; Length 1156;
Best Local Similarity 25.8%; Pred. No. 0.01;
Matches 41; Conservative 13; Mismatches 50; Indels 55; Gaps 9;

QY 154 PSP-----DWRAWQ-----RSLSLARANSQ-----DQDYXY 180
Db 46 PSPALEDEWEASETSWFWVDHFGGDDFESLAAIRFYGPVRCPLALEARTTDWAL 105

QY 181 DSTSDSNFLNPRG-W--DHTAPG-----HRTFTKQDQPEYDSTDGSDSLWSVCS 230
Db 106 PSAGERVHLNPRGFWCLNREQPRGRCSNYHVRFCPLESW-----GANGPWGPCS 159

QY 231 VTCGNGNKRTRSC----GYACTA--TESRTCDRPNCPG 263
Db 160 GSCGPGRRRLRRRCPSAGDACPCRLPQAQKCVRCPCPG 198
```

```
RESULT 9
US-08-918-914-4
; Sequence 4, Application US/08918914
; Patent No. 5876963
; GENERAL INFORMATION:
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Lawton, Michael
; APPLICANT: Magna, Holly
; APPLICANT: Yocum, Sue
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/918,914
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
```

```
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 788 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1070094
; US-08-918-914-4
```

```
Query Match 4.6%; Score 115; DB 2; Length 788;
Best Local Similarity 19.1%; Pred. No. 0.0061;
Matches 62; Conservative 36; Mismatches 120; Indels 106; Gaps 11;

QY 65 KEAPREHLHQA-----AHQFFRPRRQETGHPSLQDRPFRSFL 105
Db 230 QOREQQHREHQAQLQHQQQQQQQQQQQQPPQPQPQPQPQPSFSGTHEL 289

QY 106 DLPNFPD-----LSKADINGOXPNQVT-----IEVVDG--PDSEADKDDHPE 146
Db 290 HLQRQRQQQQQQQQQQQQQQQQQQQQQQTTFQGSQIQLOSQVPPPOOHFQQOQPQ 349

QY 147 NKPSWSVSPDWRA--WMQRLSLARANSQDQDYDYDSTSDSNFLNPRGWDHTAPGHRT 205
Db 350 QQPELERSPLDQHAQLQQQRMSQYREN-----FNQHPARPKADPCPGFC 395

QY 206 FETKQDQPEYD-----STDGEG----- 221
Db 396 APVQAPQQRPTPPVLPVINTATQPLPQPYRYPAPPACGCGCVNPPVWS 455

QY 222 ----DWSLWSVCSVTCGNGNKRTRSCG-YACTAT--ESRTCDRPNCPGIEDTFTTAATE 274
Db 456 GVNHDWSDWSTCSTCGDGAKSRRECSTNNCGADYETPCNLGPGQOTWSEWCENWSTCS 515

QY 275 VSLIAGSEEFNATKLFVVDTSCE 298
Db 516 AS--CGSGQRERTFCHLGNRCE 537
```

```
RESULT 10
US-08-313-288B-19
; Sequence 19, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```



NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
TELEX:  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1172 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-313-288B-19

Query Match 4.6%; Score 115; DB 1; Length 1172;  
Best Local Similarity 22.2%; Pred. No. 0.012;  
Matches 40; Conservative 14; Mismatches 54; Indels 72; Gaps 8;  
QY 146 ENKPSVSPDRAWQORSLARNSGDODYYDSTSDSNFLNP-----GHRFTETKDOPE 213  
DB 379 DGEEGWS-----PWAETQCSVT---CGSGTQGRGSCDVTNTCLGFSIQTRACSLSKCD 431  
QY 193 -----PRGWDHTAP-----GHTFTETKDOPE 213  
DB 432 TRIRODGGWHSWSSCVTCGVGNITRILCNSPVPMGKNGKSGRETKACQGAP- 490  
QY 214 YDSTGCE-GDWSLWSVCVTCGNGNOKRTRSC-----GYACT--ATESRTCDRPNCP 262  
DB 491 -CPIDGRWSPWSPSACTVTCAGGIRTRVCNSPEPQYGGKACVGDVQERQMCNKRSCP 549

## RESULT 11

US-08-985-526-1  
; Sequence 1, Application US/08985526  
; Patent No. 6080728  
; GENERAL INFORMATION:  
; APPLICANT: Mixson, James A  
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA  
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE  
; TITLE OF INVENTION: THERAPY  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz  
; STREET: 1220 Market Street, P.O. Box 2207  
; CITY: Wilmington  
; STATE: Delaware  
; COUNTRY: U.S.A.  
; ZIP: 19899  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/985,526  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/608,845  
; FILING DATE: 16-JUL-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McMorrow Jr., Robert G  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (302) 658-9141  
; TELEFAX: (302) 658-5613  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 218 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

## US-08-985-526-1

Query Match 4.5%; Score 113; DB 3; Length 218;  
Best Local Similarity 33.8%; Pred. No. 0.0013;  
Matches 25; Conservative 8; Mismatches 21; Indels 20; Gaps 3;  
QY 193 PRGWDHTAPGHRFTETKDOPEYDSTGEGDWSLWSVCVTCGNGNOKRTRSC---GYAC- 248  
DB 70 PRCW-----PDSADDDGWSWPSEWTSCTSTSCGNGIQGRGSCDSLNNRCE 114  
QY 249 -TATESRTCDRPNCP 261  
DB 115 GSSVQTRTCHIQEC 128

## RESULT 12

US-08-985-526-3  
; Sequence 3, Application US/08985526  
; Patent No. 6080728  
; GENERAL INFORMATION:  
; APPLICANT: Mixson, James A  
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA  
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE  
; TITLE OF INVENTION: THERAPY  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz  
; STREET: 1220 Market Street, P.O. Box 2207  
; CITY: Wilmington  
; STATE: Delaware  
; COUNTRY: U.S.A.  
; ZIP: 19899  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/985,526  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/608,845  
; FILING DATE: 16-JUL-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McMorrow Jr., Robert G  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (302) 658-9141  
; TELEFAX: (302) 658-5613  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 441 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-985-526-3

Query Match 4.5%; Score 113; DB 3; Length 441;  
Best Local Similarity 33.8%; Pred. No. 0.0039;  
Matches 25; Conservative 8; Mismatches 21; Indels 20; Gaps 3;

QY 193 PRGWDHTAPGHRFTETKDOPEYDSTGEGDWSLWSVCVTCGNGNOKRTRSC---GYAC- 248  
DB 293 PRCW-----PDSADDDGWSWPSEWTSCTSTSCGNGIQGRGSCDSLNNRCE 337  
QY 249 -TATESRTCDRPNCP 261  
DB 338 GSSVQTRTCHIQEC 351

## RESULT 13

US-08-313-288B-15

```

Query Match      4.4%; Score 111; DB 1; Length 469;
Best Local Similarity 26.1%; Pred. No. 0.007;
Matches 36; Conservative 11; Mismatches 51; Indels 40; Gaps

QY 145 PENKPSVSPDMRAWQSRSLARANSQDQDYXYDSTSDDSNFTLNPR--GWD----- 197
Db 74 PCRSRWSL-----WSTWAPCSV----- : : : : : : : : : : : : :
QY 198 HTAPGHRTFE---TKQPEYDTDGDGWSLWVCVSTCGNGNQKRTKRCGYAC----- 248
Db 114 KVAPGTLEWLQACEDQOCCPEMGCGWGPCWPPECVTSKGTTRRRACNHPAPKCGGH 173
QY 249 ---TATESRTCDRPN-CP 262
Db 174 CPGQAQSEACDTQQVCP 191

RESULT 14
US-09-341-587-3
; Sequence 3, Application US/09341587
; Patent No. 6346606
; GENERAL INFORMATION:
; APPLICANT: Mollenhauer, Jan
; TITLE OF INVENTION: Protein Containing an SRCR Domain
; FILE REFERENCE: 4121-108
; CURRENT APPLICATION NUMBER: US/09/341,587
; CURRENT FILING DATE: 1999-08-31
; EARLIER APPLICATION NUMBER: PCT/DE98/00096
; EARLIER FILING DATE: 1998-01-09
; NUMBER OF SEQ ID NOS: 12

```

```

1  ZIP: 10112
2
3  COMPUTER READABLE FORM:
4
5  MEDIUM TYPE: Floppy disk
6
7  COMPUTER: IBM PC compatible
8
9  OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: PatentIn Release #1.0, Version #1.25
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/07/862,021B
16
17 FILING DATE: 19920405
18
19 CLASSIFICATION: 435
20
21 ATTORNEY/AGENT INFORMATION:
22
23 NAME: White, John P
24
25 REGISTRATION NUMBER: 28,678
26
27 REFERENCE/DOCKET NUMBER: 40028
28
29 TELECOMMUNICATION INFORMATION:
30
31 TELEPHONE: (212) 977-9550
32
33 TELEFAX: (212) 664-0525
34
35 TELEX: 422523 COOP UI
36
37 INFORMATION FOR SEQ ID NO: 14:
38
39 SEQUENCE CHARACTERISTICS:
40
41 LENGTH: 568 amino acids
42
43 TYPE: AMINO ACID
44
45 TOPOLOGY: linear
46
47 MOLECULE TYPE: protein
48
49 US-07-862-021B-14

```

Query Match	Score	DB 1	Length
4.28	106.5	DB 1	568

Best Local Similarity 21.3%; Pred. No. 0.028;									
Matches 70;		Conservative 28;		Mismatches 122;		Indels 109;		Gaps 14;	
Qy	93	PSLQDFP	RSFL-LDLN	FPDL	SKADING	QXPNIQ	VTIE-----	VV	132
Db	129	PTIPQDK	IRPLTSLD	HPQSP	MTRG--GPII	PIARV	VIAR	KGEQCNI	IPDNVDDIVA 186
Qy	133	DGPDSE	ADKQHP	E-----	NKPSW-----	-----	SVPSPD	-----	157
Db	187	DLVTEK	EDDTP	ETC	YISN	WSPWS	ACSSAT	CDK	GKRM
Qy	158	-----	-----	-----	-----	-----	-----	-----	-----
Db	247	PCMGPG	CDD	EAST	CM	SEWIT	NSP	CS	ASCGM
Qy	199	TAPGHR	TETK	DQPEY	ST	DGEG----	DWLS	WCV	STCGN
Db	301	-----	-----	-----	-----	-----	-----	-----	-----
Qy	249	TA--TES	TC	DR	PN	CG	IED	TFR--	TAATE
Db	355	KADTE	VEK	CM	MP	ECHT	IPC	VLS	PWSE
Qy	305	SEFLK	KY	M	KY	M	N	D	L
Db	411	EELEL	K	Q	V	E	K	C	M

Search completed: April 25, 2002, 14:47:23  
Job time: 9236 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 10:06:26 ; Search time 1542.46 Seconds  
(without alignments)  
12206.639 Million cell updates/sec

Title: US-09-863-824-1  
Perfect score: 1395  
Sequence: 1 atggtgcgcctggcgccga.....aagagccagggaattataa 1395

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estnu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_hic:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_hic:\*
  - 12: gb\_gss:\*
  - 13: em\_gss\_hum:\*
  - 14: em\_gss\_inv:\*
  - 15: em\_gss\_pln:\*
  - 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	577	41.4	593	9	AW873508
2	539.2	38.7	644	10	BF472502
3	413.4	29.6	544	10	BE947736
4	407	29.2	691	10	BG162342
5	399	28.6	968	10	BF302850
6	384.2	27.5	425	9	AA884248
7	365	26.2	645	10	BJ060213
8	345.2	24.7	356	10	W86257
9	326.8	23.4	383	9	AW426875
10	284	20.4	379	9	AW495535
11	278.2	19.9	549	10	BI535390
12	260.2	18.7	479	9	AW435642
13	259.8	18.6	682	10	BF571346
14	244.4	17.5	446	10	BM191035
15	242	17.3	666	10	BI891452
16	237.8	17.0	715	10	BI704199
17	234.6	16.8	266	9	AW435639

18	233.4	16.7	1032	9	AL545675
19	211.2	15.1	643	9	BB322233
20	159.6	11.4	673	12	AO606601
21	157.8	11.3	275	10	WL4682
22	149.8	10.7	827	10	BG739780
23	147	10.5	914	12	CNS021K6
24	142.2	10.2	311	9	BB586477
25	137	9.8	679	9	AL659826
26	135.6	9.7	463	10	BG037854
27	129.4	9.3	141	9	AW435643
28	116.8	8.4	608	10	BJ071721
29	116.4	8.3	133	9	AW435648
30	93	6.7	747	9	AA447294
31	89.2	6.4	298	9	AI430783
32	81.8	5.9	693	10	BE287313
33	81.4	5.8	545	10	BI442837
34	81	5.8	589	12	AZ995663
35	80.4	5.8	353	9	AI070670
36	80.2	5.7	702	10	BJ082998
37	72	5.2	282	10	R19947
38	64.2	4.6	433	10	BI449830
39	51.4	3.7	639	10	BG344947
40	50.6	3.6	416	10	BE352595
41	50.6	3.6	595	9	AL504377
42	50.6	3.6	677	10	BE415842
43	50.6	3.6	771	12	AO861404
44	50	3.6	474	10	BM347288
45	50	3.6	506	9	AI691386

ALIGNMENTS

RESULT 1  
AW873508  
LOCUS  
DEFINITION  
ho60h09.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:3041825 3' similar to TR:095432 095432 HYPOTHETICAL 72.5 KD  
PROTEIN. ; mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AW873508 583 bp mRNA linear EST 22-MAY-2000  
ho60h09.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:3041825 3' similar to TR:095432 095432 HYPOTHETICAL 72.5 KD  
PROTEIN. ; mRNA sequence.  
AW873508  
AW873508.1 GI:8007561  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 583)  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40UP from Gibco  
High quality sequence stop: 463.  
Location/Qualifiers  
1. 583  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3041825"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH108"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NbHL19W, testis NHT, and B-cell  
NCI-CGAP-CGB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made

FEATURES  
source



```

Db 404 ATTGCCACAGAAATCCAGACCTGTGACCGCTCCAAACTGCCAGGAATTAAGATACTTTC 345
QY 805 agacagctgcaccacgaagtgaactgcttgcgggaagcagagatttaatgcacacaa 864
Db 344 AGCAGACGGCCCACTGAAGTGAAGTCTGCTTTCGGGAAGTGAAGGATTAATGCCACCAAG 285
QY 865 ctgttgaagtgcacacagacagctgtgagcgtggtgatgagctgcaaaagcgagttctta 924
Db 284 TTCTTTGAAGTTGACATGGACACCTGTGACGATGGATGAGTGCRAAAGTGAGTCTTA 225
QY 925 aagaagtataatcacaaagtgatgaatgacctgcacagctgcccctgctcctcaccact 984
Db 224 AAGAAATACATGACCAAGGTGATCAACAGCTGCCAGCTGCCAGCTGCTCTACCTACT 165
QY 985 gagtggcctacagcgcgcgcacatcttcacgcgcatcaagcgaaggaactccgcgtg 1044
Db 164 GAGGTGGCTTACAGCACAGCTGACATCTTTGACCGCATCAAGCGCAAGGACTTCCGATGG 105
QY 1045 aaggacgcagcggggcccaaggagaagctggagatctacaagcccactgcccggtactgc 1104
Db 104 AAGGATGCTAGTGGGCCCAAGAGAAACTAGAGATCTACAAAGCTACTGCTCGTACTGC 45
QY 1105 atccgctccatgctgtccctcgagagaccacacacgcgtggcgac 1147
Db 44 ATCCGCTCTATGCTGTGCTCCCTGGAGAGTACCACACTGGCGCTC 2

RESULT 3
LOCUS BE947736 544 bp mRNA linear EST 03-OCT-2000
DEFINITION UI-M-BH3-axd-h-03-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
UI-M-BH3-axd-h-03-0-UI 3', mRNA sequence.
ACCESSION BE947736
VERSION BE947736
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 544)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone Distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=NO.

FEATURES
Source
1..544
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-axd-h-03-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
```

```

ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR-amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_SEQ=None found"
BASE COUNT 146 a 165 c 100 t
ORIGIN

Query Match 29.6%; Score 413.4; DB 10; Length 544;
Best Local Similarity 85.4%; Pred. No. 1.5e-72;
Matches 472; Conservative 3; Mismatches 69; Indels 9; Gaps 1;

QY 104 cggccgcggcgaacgcagccagccagctgcagaataacctcaacgtgggaagtaca 163
Db 1 CGGCCGCGGCAACGTCAGCGGGTCCAGATTGCAGATAAATCACTCAACCTGGAGAGTACT 60

QY 164 ccacatcagaacacagcttttctctctccaaagcaccagggagcatctggaccacc 223
Db 61 CCAGTCCAGAACCCAGCTTTCTCTGTCCAAAGAACCCCGGAGGAG-----CAC 111

QY 224 aggtgcacacacacacacacacacacacacacacacacacacacacacacacacac 283
Db 112 AGGTTGTCCCAACACCTTCGCCAGACAGCGATTCCCTCCAGAGACTGGACATCCTTCAC 171

QY 284 tgcaaaagagatttcccccagatcttctctctgtatctaccacacacacacacacac 343
Db 172 TGACAGAGAGAGCGGCCCGAGATCCTTCTCTGTGACCTTACCACAACTTTCCAGATCTTCCA 231

QY 344 aagctgatataatggcagagtcacaaataccagggtcaccatagagtggtcagcagctc 403
Db 232 AAGCTGATATCAATGGCGCAAGATCCAAACATTCAGGTCAACCATAGAGTGTGTGATGGCC 291

QY 404 ctgactctgaagcagataaagatcagcatcctcgggagataaagcccagctggtcagtcacat 463
Db 292 CTGACTCCGAAGCAGAAAAAGATCAGCATCCAGAGATAAACCAGCTGGTGGCTCCCGAG 351

QY 464 cccccgactggcgccgtggtgcagaggtccctgtcttggccagggcacaacagcgagg 523
Db 352 CTCCTCGACTGCGGGCGCTGGTGGCAGAGGTCCTTGTCTTGGCCAGGACAAATAGTGGGG 411

QY 524 accaggactacragtacgacagtagtacctcagacgacagcaacttctcctcaacccccaggg 583
Db 412 ACCAGGATGACAAGTATGACAGTACCTCAGATGACAGCACTTCTCAGTCTTCTCTAGAG 471

QY 584 ggtgggaccatagacagccccagggccaccggagcttttgaacacaaagatagccaggaatg 643
Db 472 GATGGGACCGTCCAGCCCCAGGACACCGGACTTTTGAACCAACCAAGAGCAGCAGATGATG 531

QY 644 attccacagatgg 656
Db 532 ATTCCACAGACGG 544
```





Query Match	28.6%;	Score 399;	DB 10;	Length 968;	
Best Local Similarity	90.4%;	Pred. NO. 1.3e-69;			
Matches 426;	Conservative 0;	Mismatches 45;	Indels 0;	Gaps 0;	
QY 925	agaagtagcatcacaaagtgatgaatgacctgcccagctccccctgctctctacccact	984			
Db 1	AGAAATACATGCACAAGGTGATCAACGACCTGACGACCTGCCCAGCTGCTCTACCCCTACT	60			
QY 985	gaggtagctacagcacgcccagcatcttcgaccgcacaaagcgaaggaacttcgcgtgg	1044			
Db 61	GAGGTGGCTACAGCACAGCTGACATCTTTGACCGCATCAAGCCGACAGGACTTCCGATGG	120			
QY 1045	aaggacgcagcgggcccgaagaggaagctggagatgtacaaagccactgcccggtagctgc	1104			
Db 121	AAGGATGCTAGTGGGCCCAAGAGAAACTAGAGATCTTACAAGCCTACTGCTCGGTACTGC	180			
QY 1105	atccgctcatctgtccctggagagaccacgcgtggcggaagcactgctgtacagcgc	1164			
Db 181	ATCCGCTCTATGCTGTCCCTGGAGAGTACCACACTGGCTGCCACGCACTGTTGCTATGGT	240			
QY 1165	gacaacatcagctcatcacagggcgcaagggcgcgacgcgcccaacctcatcagcacc	1224			
Db 241	GACACATGCACTCATCACAGGGGCAAGGGGAGGCGACGCACTCCCATCTCATCAGCACC	300			
QY 1225	gagttctccgaggctccactacaagtgagcgtccctgcctggattatctatcgaaggg	1284			
Db 301	GAGTTCTCTGCTGAATCCACTACAAAGTGGATGTTCTGCCCCGTGATATCTGCAAGGGT	360			
QY 1285	gactggagcaggtataacgagggccgctcccaacacgacagagaagtgcacagagagc	1344			
Db 361	GACTGGAGCAGATATAATGAGGCGCGGCTCCCAATAACGGACACAGAAGTGCACAGAGC	420			
QY 1345	ccctcgagcaggaactacatacaagcagtgctccaagagccaggaattaa	1395			
Db 421	CCTTCTGATGAGACTATACATCAACAGTCTCCAAGAAGCCGAGAGTACTAA	471			
RESULT 6					
LOCUS	AA884248				
DEFINITION	am32b09.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone	425 bp	mRNA	linear	EST 04-JAN-1999
IMAGE:	1468505 3', mRNA sequence.				
ACCESSION	AA884248				
VERSION	AA884248.1	GI:2993778			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 425)				
JOURNAL	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .				
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
	Unpublished (1997)				
	Contact: Robert Strausberg, Ph.D.				
	Email: <a href="mailto:cgaps-remail.nih.gov">cgaps-remail.nih.gov</a>				
	This clone is available royalty-free through LLNL; contact the				
	IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information.				
	Insert Length: 1007	Std Error: 0.00			
	Seq primer: -40m13 fwd. Et from Amersham				
	High quality sequence stop: 424.				
FEATURES	Location/Qualifiers				
Source	1..425				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:1468505"				
	/clone_lib="Soares_NFL_T_GBC_S1"				
	/lab_host="DH10B"				
	/note="Organ: pooled; Vector: pTT3D-Pac (Pharmacia) with				
	a modified polylinker; Site_1: Not 1; Site_2: Eco RI;				
	Equal amounts of plasmid DNA from three normalized				
	libraries (fetal lung NBHL19W, testis NHT, and B-cell				

NCI-CGAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 582632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.	BASE COUNT	113 a	140 c	98 g	72 t	2 others
ORIGIN						
Query Match	27.5%;	Score 384.2;	DB 9;	Length 425;		
Best Local Similarity	96.9%;	Pred. NO. 9.5e-67;				
Matches 409;	Conservative 3;	Mismatches 7;	Indels 3;	Gaps 2;		
QY 188	ttctccaaaagaagcaccacaaaggagcatctggaccaccaggtgtgcacacaccccttccccca	247				
Db 1	TTCTCCAAAGAAGCACCAAGGGAGCATCTGGACCCACGAGCTGCACCAACCCCTTCCCCA	60				
QY 248	gaccgcgattccgacaagagacggggcaccccttcattgcaagagatcttccccagatcct	307				
Db 61	GACCGCGATTCCGACAAAGAGACGGGGCA-CCTTCATTGCAAGAGATTTCCCCAGATCCT	119				
QY 308	ttctcttgatctaccaaaactttccagatctttccaaagctgatatcaatgggcgrwtc	367				
Db 120	TTCTCTTGATCTACCAAACTTTCCAGATCTTTCCAAAGCTGATATCAATGGCGAATC	179				
QY 368	caaatccaggtccaccatagagggtggtgcagcgttcctgactctgaacagacataaagatc	427				
Db 180	CAAATATCCAGGTTCACATAGAGGTGGTCGACGGTCTCTGACTCTGAAGCAGATAAAGATC	239				
QY 428	agcatccgagataagcccagctggtcagtcctccctcccagctggtggtggtggtggtg	487				
Db 240	AGCATCCGAGATATAAGCCACAGCTGGTCACTCCCATCCCCGACTGGCGGGCTGGTGGC	299				
QY 488	agaggtccctgtccttggccagggcgaacagcggggacagcagctacragtacgacagta	547				
Db 300	AGAGGTCCCTGTCTCTTGGCCAGGGCAACAGCGGGGACGAGACTACAAGTACGACAGTA	359				
QY 548	ctctcagacagcagaactctctcaacccccccaggg--ggtgggaccatacagccccaggg	605				
Db 360						



```

Db 301 GCACTGCTGCTAGCGGACAAACATGCAGCTCATCACAGGNCAGGNGTCGGCA 356

RESULT 9
AW426875
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 383)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Cassas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perlea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCGCCAGTCAGCAGC
Plate: 26 row: L column: 24
Seq primer: ATTAGGTGACACTATAG.

FEATURES
source
Location/Qualifiers
1..383
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT 97 a 104 c 113 g 69 t
ORIGIN

Query Match 23.4%; Score 326.8; DB 9; Length 383;
Best Local Similarity 92.7%; Pred. No. 2.5e-55;
Matches 343; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 697 tcgaggacgcgcacagacagaaacgagaccggtctgtggtacgcgtgcacgcaacagaa 756
Db 14 TGTGGGAATGGCAACCAACGAGACGACGAGCTTGTGGCTACGCTACTGCAACCGAA 73
QY 757 tcagagacctgtaccgtccaaactcccaggaattgaagacacacttttagacagctgcc 816
Db 74 TCTAGGACGTGTGACCGTCCCGCTGCGGGAATGAAGACACCTTTCAGGACAGCTGCC 133
QY 817 accgaagtgtctgtcttcggaagcagagagtttaatgccacaaactgtttgaagtt 876
Db 134 ACTGAAGTGTGCTGCTTGGCGGAAGTGAAGAGTTTAATCCCAACCACTTTTGAAGTC 193
QY 877 gacacagacagctgtgagcgtggtgatgagctgcaaaagcaggttcttaagaagatcatg 936
Db 194 GACACGACACGCTGTGAGCGCTGGATGAGTTGCAAGAGCGAGTTCCTTAAGAAATACATG 253

```

```

QY 937 cacaagtgatgaatgacctgccagctgccctgtcctaccctcaggtgagtgacctac 996
Db 254 CATAAGGTGATCAACGACCTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 313
QY 997 agcacggccgacatcttcgaccgcacatcaagcgaagacgttcgctggaagagcgccagc 1056
Db 314 AGCACAGCTGACATCTTCGACCGCATCAAGCGCAAGGACTTCGCTGCGAAGGAGCCGCG 373
QY 1057 gggcccaagg 1066
Db 374 GGGCCCAAGG 383

RESULT 10
AW495535
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AW495535 379 bp mRNA linear EST 24-FEB-2000
UI-M-BH3-avb-e-02-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
UI-M-BH3-avb-e-02-0-UI 3', mRNA sequence.
AW495535
AW495535.1 GI:7065816
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 379)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL
MEDLINE
COMMENT

Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
Oligo-dT track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENETICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=No.

FEATURES
source
Location/Qualifiers
1..379
/organism="Mus musculus"
/db_xref="C57BL/6J"
/clone="UI-M-BH3-avb-e-02-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived

```



JOURNAL  
COMMENT

Unpublished (2000)  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -mnscore 20  
and -minmatch 12 options.

## PCR Primers

FORWARD: AGGAACACGTATGACCAT

BACKWARD: GTTTCACGATCAGCAGC

Plate: 24 row: L column: 20

Seq primer: ATTTAGTGACACTATAG.

FEATURES  
source

1..682  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4252124"  
/clone\_lib="NIH\_MGC\_62"  
/tissue\_type="melanotic melanoma, high MDR"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: skin; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggcgctcgcc); Site\_2: SfiI (ggccattatggcc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCGCGCGCGGCACATG-dT(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."

BASE COUNT 171 a 183 c 212 g 116 t

## ORIGIN

Query Match 18.7%; Score 260.2; DB 9; Length 479;  
Best Local Similarity 93.4%; Pred. No. 5e-42;  
Matches 268; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

QY 359 ggcagrvtcacaaatccaggctcaccatagaggtggtcgacggctcctgactctgaagcag 418  
Db 12 GGCAGATCCAAATATCCAGGTCACCATAGAGTGGTGACGGTCTGAAATGTGAAGCAG 71  
QY 419 ataaagatcagatcccgagagaataagccagctggtcagtcctccatcccccgaactggcggg 478  
Db 72 ATAAAGATCAGATCCCGAGAGTAAGCCAGCTGCTCAGTCCCATCCCTGACTTGGCGGG 131  
QY 479 cctggtggcagaggtcctctgcttggccaggggcaaacagcggggagcaggaactacragt 538  
Db 132 CCTGGTGGCAGAGTCTCTGTCTTGGCAAGAGCCACAGCGGGGACGAGACTTACAAGT 191  
QY 539 acgacgtactcagacagcagaactctcctcaacccccccagggggtggaccatacag 598  
Db 192 ACGACGTACTCAGATGACAGCAACTTCTCTCAACCCCCCGGGGGCTGGGACCATCCGG 251  
QY 599 cccagggccacgggacttttgaacacaaagatcagcagaatatgat 645  
Db 252 CCCAGGCCACCGGACTTTCGAAACCAAGAGAGCAGCAGCAATATGGT 298

## RESULT 13

BF571346 682 bp mRNA linear EST 12-DEC-2000  
LOCUS 602077485F1 NIH\_MGC\_62 Homo sapiens cDNA clone IMAGE:4252124 5',  
DEFINITION mRNA sequence.

ACCESSION BF571346

VERSION BF571346.1 GI:11645058

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

## REFERENCE

1 (bases 1 to 682)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC/DCTD/DTF

CDNA Library Preparation: CLONETECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM1076 row: i column: 21  
High quality sequence stop: 379.

FEATURES  
source

1..682  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4252124"  
/clone\_lib="NIH\_MGC\_62"  
/tissue\_type="melanotic melanoma, high MDR"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: skin; Vector: pDNR-LIB (Clontech); Site\_1:  
SfiI (ggcgctcgcc); Site\_2: SfiI (ggccattatggcc);  
Double-stranded cDNA was prepared from cell line RNA. 5'  
and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCGCGCGCGGCACATG-dT(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."

BASE COUNT 171 a 183 c 212 g 116 t

## ORIGIN

Query Match 18.6%; Score 259.8; DB 10; Length 682;  
Best Local Similarity 88.8%; Pred. No. 6.4e-42;  
Matches 420; Conservative 0; Mismatches 32; Indels 21; Gaps 12;

QY 939 caagtgatgaatgacgtcccgagctgcccctgctctaccctactgagtggtgctacag 998  
Db 1 CAAGGTGATGAATGACCTGCCAGCTGCCCTGCTCTACCCCACTGAGTGGCTACAG 60  
QY 999 cagcgccgacatcttcgacgcgcatcaagcgcaaggaactccgtggaggagcgacgagg 1058  
Db 61 CACGG-CGACATCTTCGACCGCATCAAGCGCAAGGACTTCGCGCTGGAAGGACGCCAGCGG 119  
QY 1059 gccca-agggagaagctggagatctacaagccactgccggtactgctatcgctccatgc 1117  
Db 120 GCCCATAGGAGAAGCTGGAGATCTACAAGCCCACTGCCCGTACTGCTATCCG-TCCATGC 178  
QY 1118 tgtccctggagagcaccacgcctggcgccagcactgctgctacggtgacacatgcaagc 1177  
Db 179 TGTCCCTGGAGAGCACCCAGCTGGCGGCACAGAACTGTTG--TACGGGACAAATGCAGC 236  
QY 1178 tcataccaggggcaagggggcgccagcgcaccccaacctcatcagcaccgagttctccgag 1237  
Db 237 TCATCACCGGGGCAAGGGGGGGGAAAG-CCAAACTCATCAGCACCAGTTCCTCCGCGG 295  
QY 1238 agctccact--acaaggtggacgtcctgccc-tggattatctg---caaggggtgactgga 1291  
Db 296 AGCTCCAATTACAAGGTGGAGCTCTCCCGTGGATTATCTGCAAAAGGTGTGACTGGA 355  
QY 1292 g-caggtataacgagggcccgccctcccaacaacg-----gacggaagtgccagagagc 1344  
Db 356 GCCAGGTATACGAGGCGCGGCTCCACAAAGCGGGAACAGACGCTGGGCCAGAGAGC 415  
QY 1345 ccctcggac-gagacatcatcaagc-agttccaaagggccagggaattataa 1395  
Db 416 CCCTCGGACGGAGGACTACATCAAGCAAGTTTCCAAGAGGCCAGGACATTTTAA 468

## RESULT 14

BM191035 446 bp mRNA linear EST 11-DEC-2001  
LOCUS BM191035  
DEFINITION 5' similar to TR:095432 095432 HYPOTHETICAL 72.5 KD PROTEIN. ;

1

```
Db 236 TCAAGCCACTCGGAGACTGGGAGAGAAGAGTGAAGAGTGAAGCTGGAGCGGGAAGCAGAA 295
Qy 629 atcagccagaatatgattccacagatggcgaggtgactgagtgctctctggtctctgtctgca 688
Db 296 CCCAGACTGAATACGACTACATACCGGGGAAGCGGACTGGAGTGTGGTCTCCATGCA 355
Qy 689 gcgtacctgcgggaacgggaaccagaaacggaccgccggtcttcttggtctacgcgtgcactg 748
Db 356 GTGTATCTCTGTGGAATGGTAAACGAGAAGCGGACAGATCGTGGCGTTATGCCCTGCACTG 415
Qy 749 caacagaatcgagagacctgtgaccgtccaaactgccaggaaatgaagacaccttttagga 808
Db 416 CCACAGAGTCAACGACATGTGACATGCCCACTGTCCAGGGATTGAAGATGCATTCAAGA 475
Qy 809 cagctgccaccgaagtgcgtcttcggaagcgaggagtttaatgccaccacaaactgt 868
Db 476 CGGCAGCACTGAAGTTCAGTTTGCTCGCTGGCACTGAAGAGTCAATGGCACAGAGCTGT 535
Qy 869 ttgaagttgcacacagactgtgagcgtgagcgtgagcgtgcaaaagcgagttcttaaaga 928
Db 536 TTGGNAGTGATACCGACAGCTGTGAGCGGTGGATGAAGTGAAGAGTGAAGTGTCTCAAGA 595
Qy 929 agtcatgcacaaggtgatgaatgacctgcccagctgcccctgctctctacccccactgagg 988
Db 596 A-TACATGAGTAAAGTAGCCACTGATCTACCCAGCTGCCCTGCTTTTACCCACCGAGG 654
Qy 989 tggcctacagca 1000
Db 655 TGGNCTACAGCA 666
```

Search completed: April 25, 2002, 11:39:10  
Job time: 5564 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 11:11:52 ; Search time 204.53 Seconds  
(without alignments)  
11710.241 Million cell updates/sec

Title: US-09-863-824-1

Perfect score: 1395

Sequence: 1 atggcgccctggcgccga.....aagagccagggaattattaa 1395

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1393.8	99.9	1395	24	Human thrombospondin
2	875.8	62.8	936	24	Human thrombospondin
3	738.6	52.9	739	22	Human protein enco
4	554.6	39.8	790	21	Human OREF1686
5	495	35.5	495	24	Human thrombospondin
6	447.6	32.1	448	23	DNA encoding novel
7	325.2	23.3	2981	22	Human cDNA encoding
8	325.2	23.3	2981	22	Human proliferatio
9	228.4	16.4	790	21	Human OREF1686

Key CDS Location/Qualifiers  
1..1395  
/\*tag= a  
/product= "Thrombospondin-like protein"  
/transl\_except= (pos:364..366, aa:Xaa)  
/note= "Xaa is Asn or Val"  
/transl\_except= (pos:535..537, aa:Xaa)  
/note= "Xaa is Lys or Glu"  
364  
misc\_feature  
/\*tag= b  
/note= "This degenerate base represents A-G transition polymorphism"  
365  
misc\_feature  
/\*tag= c  
/note= "This degenerate base represents A-T transversion polymorphism"

## ALIGNMENTS

RESULT 1

AAD23967

ID AAD23967 standard; cDNA; 1395 BP.

XX AAD23967;

XX 26-MAR-2002 (first entry)

XX Human thrombospondin-like protein cDNA #1.

XX Human; thrombospondin-like protein; therapeutic; pharmacogenomic;  
XX drug screening; cosmetic; nutraceutical; anti-angiogenic; wound healing;  
XX endocrine; gene therapy; hyperthyroidism; hypothyroidism; ss.

OS Homo sapiens.

XX Key CDS Location/Qualifiers

1..1395  
/\*tag= a  
/product= "Thrombospondin-like protein"  
/transl\_except= (pos:364..366, aa:Xaa)  
/note= "Xaa is Asn or Val"  
/transl\_except= (pos:535..537, aa:Xaa)  
/note= "Xaa is Lys or Glu"  
364  
misc\_feature  
/\*tag= b  
/note= "This degenerate base represents A-G transition polymorphism"  
365  
misc\_feature  
/\*tag= c  
/note= "This degenerate base represents A-T transversion polymorphism"







KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200058473-A2.  
 PN  
 XX  
 XX 05-OCT-2000.  
 PD  
 XX  
 XX 31-MAR-2000; 2000WO-US08621.  
 XX  
 XX 31-MAR-1999; 99US-0127607.  
 PR  
 XX 02-APR-1999; 99US-0127636.  
 PR  
 XX 05-APR-1999; 99US-0127728.  
 PR  
 XX 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 XX Shinkets RA, Leach M;  
 PI  
 XX WPI; 2000-602362/57.  
 DR  
 XX P-PSDB; AAB41922.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 PT  
 XX  
 PS Claim 5; Page 2553; 5507pp; English.  
 XX  
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 XX which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy.  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 790 BP; 181 A; 238 C; 212 G; 157 T; 2 other;

Query Match 39.8%; Score 554.6; DB 21; Length 790;  
 Best Local Similarity 98.2%; Pred. NO. 1.5e-115;  
 Matches 557; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
 QY 176 ccagcttttctctccaaagaagaccagggagcatctggaccacaggtgcacacc 235  
 DB 224 ccgcttttgccttggccaaagaagaccagggagcatctggaccacaggtgcacacc 283

QY	236	aaccttccccagacgcgattccgacaagagacgggccccttctcattgcaagagatt	295
DB	284	acaccttccccagacgcgattccgacaagagacgggccccttctcattgcaagagatt	343
QY	296	tcctccagatcttctctctgattctaccacaaacttccagatcttccaaagctgatca	355
DB	344	tcctccagatcttctctctgattctaccacaaacttccagatcttccaaagctgatca	403
QY	356	atgggcagrvtcacaaatccaggtccaccatagaggtggtcgacggctcctgactctgaag	415
DB	404	atgggcagrvtcacaaatccaggtccaccatagaggtggtcgacggctcctgactctgaag	463
QY	416	cagataaagatcagatcccggaataagcccagctggttcagtcctccatccccgactggc	475
DB	464	cagataaagatcagatcccggaataagcccagctggttcagtcctccatccccgactggc	523
QY	476	gggccttggtggcagaggtccctgtccttggccaggggcaaacagcgggggacagactacr	535
DB	524	gggccttggtggcagaggtccctgtccttggccaggggcaaacagcgggggacagactaca	583
QY	536	agtaacagatcactcagacgacgaacttctcctcaaccccccagggggtggaccata	595
DB	584	agtaacagatcactcagacgacgaacttctcctcaaccccccagggggtggaccata	643
QY	596	cagccccagggccacccgacttttgaacacaaagatcagcagaatattgattccacagatg	655
DB	644	cagccccagggccacccgacttttgaacacaaagatcagcagaatattgattccacagatg	703
QY	656	gcaggggtgactgaggtctctgtctgtctgcagcgtccactgcgggaacggaacacaga	715
DB	704	gcaggggtgactgaggtctctgtctgtctgcagcgtccactgcgggaacggaacacaga	763
QY	716	aacggacccggtctgtggtacgcgt 742	
DB	764	aacggacccggtctgtggtacgcgt 790	
RESULT 5			
ID	AAD23968		
XX	AAD23968 standard; CDNA; 495 BP.		
XX	AAD23968;		
XX	26-MAR-2002 (first entry)		
DE	Human thrombospondin-like protein cdna #2.		
XX	Human; thrombospondin-like protein; therapeutic; pharmacogenomic;		
KW	drug screening; cosmetic; nutraceutical; anti-angiogenic; wound healing;		
KW	endocrine; gene therapy; hyperthyroidism; hypothyroidism; ss.		
XX	Homo sapiens.		
XX	Key	Location/Qualifiers	
FT	CDS	1..495	
FT	/*tag= a		
FT	/product= "Thrombospondin-like protein"		
XX	WO200190179-A2.		
XX	29-NOV-2001.		
XX	23-MAY-2001; 2001WO-US16786.		
XX	23-MAY-2000; 2000US-206415P.		
XX	(LEXI-) LEXICON GENETICS INC.		
XX	Turner CA, Hilbun E, Donoho G, Friedrich G, Abuin A, Zambrowicz B;		
XX	Sands AT;		
XX	WPI; 2002-089923/12.		
DR	P-PSDB; AAE14398.		

XX New isolated nucleic acid molecule that encodes human  
PT thrombospondin-like protein, useful in therapeutic, diagnostic and  
PT pharmacogenomic applications, in drug screening and clinical trial  
PT monitoring  
XX  
XX Claim 1: Page 37; 39pp; English.  
XX  
XX The invention relates to nucleic acid molecule encoding human  
CC thrombospondin-like protein. The polynucleotide and polypeptide of  
CC the invention are useful in therapeutic, diagnostic and pharmacogenomic  
CC applications, in drug screening, clinical trial monitoring, and in  
CC cosmetic or nutraceutical applications. The polynucleotide is useful for  
CC identifying coding sequences and mapping a unique gene to a particular  
CC chromosome, to screen libraries, isolate clones, prepare cloning and  
CC sequencing templates, as hybridisation probes for screening libraries  
CC and assessing gene expression patterns, in microarrays and other assay  
CC formats to screen collections of genetic material from patients who  
CC have a particular medical condition, to identify mutations associated  
CC with a particular disease, in a diagnostic or prognostic assay, as  
CC antisense molecules, and as part of ribozymes and/or triple helix  
CC sequences that are useful for gene regulation, for detecting mutant  
CC proteins or inappropriately expressed proteins for the diagnosis of  
CC disease and in the molecular mutagenesis/evolution of proteins that  
CC are at least partially encoded by the polynucleotide. The  
CC thrombospondin-like protein or its modified processed form can  
CC be used as therapeutics e.g. anti-angiogenic agents, to promote wound  
CC healing and regulate endocrine functions. The polynucleotide  
CC can also be used in gene therapy and for treating disorders like  
CC hyperthyroidism and hypothyroidism. The present sequence is a cDNA  
CC encoding human thrombospondin-like protein.  
XX  
XX Sequence 495 BP; 126 A; 156 C; 139 G; 74 T; 0 other;

Query Match 35.5%; Score 495; DB 24; Length 495;  
Best Local Similarity 100.0%; Pred. No. 3.6e-102;  
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 901 atgagctcaaaagcaggtctttaaagatcatatgcacaggtgatgaatgacctgcc 960  
Db 1 atgagctcaaaagcaggtctttaaagatcatatgcacaggtgatgaatgacctgcc 60  
Qy 961 agctgccccctgctctacccactgagtggtgctcctacagcagcgccgacatttcgacgc 1020  
Db 61 agctgccccctgctctacccactgagtggtgctcctacagcagcgccgacatttcgacgc 120  
Qy 1021 atcaagcgaagcagcttcctggaagacagccagcagcgccgacaggaagctgagatc 1080  
Db 121 atcaagcgaagcagcttcctggaagacagccagcagcgccgacaggaagctgagatc 180  
Qy 1081 tacaagccactgcccgtactgctcctgctccatgctgctcctggagagcaccacgctg 1140  
Db 181 tacaagccactgcccgtactgctcctgctccatgctgctcctggagagcaccacgctg 240  
Qy 1141 gcggcagcagcagctgtctcagcgagacacacacacacacacacacacacacacac 1200  
Db 241 gcggcagcagcagctgtctcagcgagacacacacacacacacacacacacacacac 300  
Qy 1201 ggcagcgcac 1260  
Db 301 ggcagcgcac 360  
Qy 1261 ctgcccctgattatctcaagggtgactggagcaggtataacgagggcccgctcccaac 1320  
Db 361 ctgcccctgattatctcaagggtgactggagcaggtataacgagggcccgctcccaac 420  
Qy 1321 aacgagcagaagtgcagagagacccctcgagcagaggtacatcaacagcttccaag 1380  
Db 421 aacgagcagaagtgcagagagacccctcgagcagaggtacatcaacagcttccaag 480  
Qy 1381 gccaggggaattaa 1395  
|||||

Db 481 gccaggggaattaa 495  
RESULT 6  
AAS70533  
ID AAS70533 standard; cDNA; 448 BP.  
XX  
XX AAS70533;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX DNA encoding novel human diagnostic protein #6337.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX P-PSDB; ABG06346.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX Claim 1; SEQ ID NO 6337; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 448 BP; 123 A; 122 C; 123 G; 80 T; 0 other;

Query Match 32.1%; Score 447.6; DB 23; Length 448;  
Best Local Similarity 99.8%; Pred. No. 1.7e-91;  
Matches 447; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 522 ggaccagactacagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 581  
|||||  
Db 1 ggaccagactacagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 60  
|||||









PR 05-APR-1999; 990US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX (CURA-) CURAGEN CORP.  
PA Shinkets RA, Leach M;  
PI WPI; 2000-602362/57.  
DR P-PSDB; AAB41922.  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX Claim 5; Page 2553; 5507pp; English.  
XX AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianaemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
XX Sequence 790 BP; 181 A; 238 C; 212 G; 157 T; 2 other;

Query Match 16.4%; Score 228.4; DB 21; Length 790;  
Best Local Similarity 91.3%; Pred. No. 4.8e-42;  
Matches 252; Conservative 1; Mismatches 22; Indels 1; Gaps 1;

QY 467 ccgactgaggcggctggtgcagaggtccctgtccttggccaggggcaaacagggggacc 526  
DB 277 CAGCCTGGTGGTCCAGATGCTCCCTTGGTCTTCTTTGGCCAGGGGAAA-AGCGGGGACC 219  
QY 527 aggactacragtacgacgtacctcagacgacgacgaacttccctcaacccccccaggggggt 586  
DB 218 AGGACTACAGTACGGCAGTACCTCAGACGACGACACTTCCCTCAACCCCCCAGGGGGT 159  
QY 587 ggagaccatacagcccccagggccacgggacttttgaacccaagaatcagccagaatatgatt 646  
DB 158 GGGACCATACAGGCCCCAGGGCCAGCGGACTTTTGAACCAAGAATCAGCCGAATGATT 99  
QY 647 ccacaatgcgaggtgactgagctctgtgtctgtcgtcagcgtcactgcgggaacg 706  
DB 98 CCACAGATGGCGAGGTGACTGGAGTCTCTGGTCTGTCTGCAGCGTCACTGCGGGAACG 39  
QY 707 gcaaccgaacacgacccggctcttgtgtgctacgcgt 742  
DB 38 GCAACCAAGAACGACCCCGCTCTGTGCTACGGT 3

RESULT 10  
ID AAS70534  
XX AAS70534 standard; cDNA; 648 BP.  
AC AAS70534;  
XX  
XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #6338.  
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
PI WPI: 2001-639362/73.  
DR P-PSDB; ABG06347.  
DR New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 1; SEQ ID No 6338; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 648 BP; 201 A; 155 C; 163 G; 129 T; 0 other;

Query Match 10.4%; Score 144.6; DB 23; Length 648;  
Best Local Similarity 97.4%; Pred. No. 3.4e-23;  
Matches 147; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 637 gaatatgattccacagatgaggggagtgactggagtgctctgtgtcgtcagcgtcacc 696  
DB 58 gaggaataattccacagatgaggggagtgactggagtgctctgtgtcgtcagcgtcacc 117  
QY 697 tgcgggaacggaaccagaaacgacccggtcttctgtgtgtacggtcactgcaacagaa 756  
DB 118 tgcgggaacggaaccagaaacgacccggtcttctgtgtgtacggtcactgcaacagaa 177  
QY 757 tcagagacctgtaccgtccaaactgccacg 787  
DB 178 tcagagacctgtaccgtccaaactgccacg 208



[illegible]

Db	151	agcgacttctaatacagtatctgagccagatgctgcgggaacctgcacagntgcccg	210
Qy	972	ctcctacccccactgagtgccctacagcagccgcacatcttcacgcgcatcaacgcgcaa	1031
Db	211	tgctaccacctggagggccatggagacgacctgagcctcagggagagaccagggcg	270
Qy	1032	ggacttcgctgggaagacgcagcgggcccaagagagaagctggagatctacaagccca	1091
Db	271	cagcttcggtggagggatgccagtgccctgcgcagcgctggacatctaccagccca	330
Qy	1092	tgcccggtactgcactccgttcaatgctgtctccttgagagagaccacagctggcgggcacagca	1151
Db	331	ggcgcgcttctgcctcgttccatgctgtctgtgggagagacacagctgcgc-cagca	389
Qy	1152	ctctgtctagcgcgacaacatcagc-tcatcaccaggggcaaggggcgagcagccca	1210
Db	390	ctgcgtctatgcagggacagcgcgcttctgacctgtgcgaaggcgccggcatgcca	449
Qy	1211	acctcatca	1219
Db	450	ctatnagna	458

RESULT	13	
AAI178722		
ID	AAI178722	standard; DNA; 51 BP.
XX		
AC	AAI178722;	
XX		
DT	09-NOV-2001	(first entry)
XX		
DE	Human silent SNP containing nucleic acid SEQ:5663.	
XX		
XX	Human; single nucleotide polymorphism; SNP; genome; gene therapy;	
KW	protein therapy; vaccine; probe; diagnostic assay; detection;	
KW	quantitation; restorative therapy; polymorphic; ds.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200140521-A2.	
XX		
PD	07-JUN-2001.	
XX		
PF	30-NOV-2000; 2000WO-US32758.	
XX		
PR	30-NOV-1999; 99US-0168138.	
PR	29-NOV-2000; 2000US-0726173.	
XX		
PA	(CURA-) CURAGEN CORP.	
XX		
PI	Shimkets RA, Leach M;	
XX		
XX	WPI; 2001-356160/37.	
DR		
XX		
PT	Polymorphic nucleic acid sequences, useful in genetic testing and	
PT	therapy -	
XX		
PS	Claim 1; Page 2243; 2653pp; English.	

CC in samples, and therefore which patients may be in need of restorative  
CC therapy. The polypeptides encoded by (I) may be used as antigens in the  
CC production of antibodies specific for polymorphic polypeptides. The  
CC antibodies may also be used to down regulate expression and activity.  
CC The antibodies may also be used as diagnostic agents for detecting the  
CC presence of polymorphic polypeptides in samples.  
XX  
XX Sequence 51 BP; 17 A; 19 C; 11 G; 4 T; 0 other;

Query Match	3.7%	Score 51	DB 22	Length 51
Best Local Similarity	100.0%	Pred. No.	0.021	
Matches	51	Conservative	0	Mismatches 0
		Indels	0	Gaps 0

Qy 191 ccaaaagcaccagggagcatctggaccaccaggtgcacacccct 241  
 bh 1 ccaaaagcaccaggaagagcatctgaccaccaggtgcacacccct 51

RESULT	14
AAI78724	
ID	AAI78724 standard; DNA; 51 BP.
XX	
AC	AAI78724;
XX	
DT	09-NOV-2001 (first entry)
XX	
DE	Human silent SNP containing nucleic acid SEQ:5665.
XX	
KW	Human; single nucleotide polymorphism; SNP; genome; gene therapy;
KW	protein therapy; vaccine; probe; diagnostic assay; detection;
KW	quantitation; restorative therapy; polymorphic; ds.
XX	
OS	Homo sapiens.
XX	
PN	W0200140521-A2.
XX	
PD	07-JUN-2001.
XX	
PF	30-NOV-2000; 2000WO-US32758.
XX	
PR	30-NOV-1999; 99US-0168138.
PR	29-NOV-2000; 2000US-0726173.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shimkets RA, Leach M;
XX	
DR	WPI; 2001-356160/37.
XX	
PT	Polymorphic nucleic acid sequences, useful in genetic testing and
PT	therapy -
XX	
PS	Claim 1; Page 2244; 2653pp; English.

AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide sequences (I), which contain single nucleotide polymorphisms (SNPs). AAI53114 to AAI53329 represent peptides related to human polymorphic polynucleotide sequences. The sequences can be used in gene and protein therapy, and in vaccine production. (I) and the polypeptides encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate expression of polymorphic polypeptides. For example, (I) may be used to treat disorders by rectifying mutations or deletions in a patient's genome that affect the activity of polypeptides by expressing inactive proteins or to supplement the patients own production of polypeptide. Additionally, (I) and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. The polypeptides encoded by (I) may be used as antigens in the production of antibodies specific for polymorphic polypeptides. The antibodies may also be used to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2002, 10:07:51 ; Search time 1729.22 Seconds  
(without alignments)  
16881.900 Million cell updates/sec

Title: US-09-863-824-1  
Perfect score: 1395  
Sequence: 1 atgggtgccttgcggccga.....aagagggccagggaattataa 1395

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: gb.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
-----					

1	1393.8	99.9	1395	6	AX351002	AX351002 Sequence
2	875.8	62.8	936	6	AX351006	AX351006 Sequence
3	519	37.2	84122	9	AL133463	AL133463 Human DNA
4	495	35.5	495	6	AX351004	AX351004 Sequence
5	457	32.8	1617	9	BC017997	BC017997 Homo sapi
6	326.8	23.4	3052	9	AK056709	AK056709 Homo sapi
7	325.2	23.3	2981	6	AX136231	AX136231 Sequence
8	268.8	19.3	231464	9	AF111168	AF111168 Homo sapi
9	183.2	13.1	107057	9	HSJ10712	AL050320 Human DNA
10	164.2	11.8	523	11	HS107712S	AL110084 H. sapiens
11	55.2	4.0	125020	9	AF429315	AF429315 Homo sapi
12	54.2	3.9	45519	1	FVBPOAD2A	D26094 Flavobacter
13	52.8	3.8	1347	9	AF257167	AF257167 Homo sapi
14	52.4	3.8	7585	1	SCY14206	Y14206 Streptomyce
15	52.4	3.8	42816	1	SCC88	AL139298 Streptomy
16	52.2	3.7	34182	1	SC111	AL096849 Streptomy
17	51.8	3.7	42210	1	SC1C2	AL031124 Streptomy
18	51	3.7	51	6	AX162335	AX162335 Sequence
19	51	3.7	51	6	AX162337	AX162337 Sequence
20	50.6	3.6	3778	1	AVIALGEB	L39013 Azotobacter
21	50.6	3.6	15759	1	AVIALGEB	L39096 Azotobacter
22	50.2	3.6	51	6	AX162339	AX162339 Sequence
23	50.2	3.6	51	6	AX162341	AX162341 Sequence
24	50.2	3.6	1839	6	AX196078	AX196078 Sequence
25	50.2	3.6	11905	1	APU33059	U33059 Actinosynne
26	50.2	3.6	109519	6	AX195929	AX195929 Sequence
27	50.2	3.6	161970	8	AP003221	AP003221 Oryza sat
28	50	3.6	37245	1	SC5F2A	AL049587 Streptomy
29	49.8	3.6	6040	1	CTU32622	U32622 Comamonas t
30	49.6	3.6	3108	9	HUMIMUCB	M94132 Human mucin
31	49.6	3.6	15720	6	AX193491	AX193491 Sequence
32	49.6	3.6	15720	6	AX330186	AX330186 Sequence
33	49.6	3.6	15720	9	HUMMUC2X	L21998 Homo sapien
34	49.4	3.5	10959	1	AE004953	AE004953 Pseudomon
35	49.4	3.5	125020	9	AF429315	AF429315 Homo sapi
36	49	3.5	1149	6	AX089208	AX089208 Sequence
37	48.8	3.5	1158	6	AR111753	AR111753 Sequence
38	48.8	3.5	1158	6	AX089125	AX089125 Sequence
39	48.8	3.5	152396	8	AP002971	AP002971 Oryza sat
40	48.6	3.5	51	6	AX162340	AX162340 Sequence
41	48.6	3.5	51	6	AX162342	AX162342 Sequence
42	48.4	3.5	592	8	WHTPH288B	D37943 Wheat mRNA
43	48.4	3.5	158186	2	AC096212	AC096212 Rattus no
44	48.2	3.5	137332	8	AP002820	AP002820 Oryza sat
45	48	3.4	2888	6	A48323	A48323 Sequence 1

## ALIGNMENTS

RESULT 1	AX351002	AX351002	Sequence	1395 bp	DNA	linear	PAT 06-FEB-2002
LOCUS	AX351002	Sequence	1 from Patent WO0190179.				
ACCESSION	AX351002						
VERSION	AX351002.1	GI:18616372					
KEYWORDS	human.						
SOURCE	Homo sapiens						
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.						
REFERENCE	1 (sites)						
AUTHORS	Turner,C.A., Hilbun,E., Donoho,G., Friedrich,G.C., Abuin,A., zambrowicz,B. and Sands,A.T.						
TITLE	Novel human thrombospondin-like proteins and polynucleotides encoding the same						
JOURNAL	Patent: WO 0190179-A 1 29-NOV-2001;						
FEATURES	Lexicon Genetics Incorporated (US)						
source	Location/Qualifiers						
	1..1395						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
BASE COUNT	345 a	431 c	232 t				3 others







repeat_region	2642.. 2951	/note="AluJo repeat: matches 1. .311 of consensus"
repeat_region	3460.. 3507	/note="24 copies 2 mer tg 79% conserved"
repeat_region	4065.. 4106	/note="21 copies 2 mer cc 76% conserved"
misc_feature	complement(442..5302)	/note="match: GSS: Em:AQ741486"
repeat_region	4908.. 5154	/note="MIR repeat: matches 34.. .238 of consensus"
misc_feature	5180.. 5672	/gene="ba149i18.1"
repeat_region	7941.. 8067	/note="match: GSS: Em:AQ755240"
repeat_region	8078.. 8510	/note="MIR repeat: matches 1. .140 of consensus"
repeat_region	8538.. 8624	/note="MLT1H repeat: matches 81. .519 of consensus"
misc_feature	complement(8621..9044)	/note="MIR repeat: matches 170. .262 of consensus"
misc_feature	complement(8758..9044)	/note="match: GSS: Em:AQ182554"
repeat_region	9887.. 10251	/note="match: GSS: Em:AQ133736"
repeat_region	10536.. 10673	/note="MLT1I repeat: matches 19. .371 of consensus"
repeat_region	10826.. 11052	/note="L2 repeat: matches 2563.. .2708 of consensus"
repeat_region	11277.. 11343	/note="MIR repeat: matches 2. .261 of consensus"
repeat_region	11344.. 11517	/note="MERSA repeat: matches 99. .166 of consensus"
repeat_region	11520.. 11698	/note="MERSA repeat: matches 1. .182 of consensus"
repeat_region	11773.. 11825	/note="MERSB repeat: matches 2. .178 of consensus"
repeat_region	12156.. 12212	/note="MERSA repeat: matches 136. .187 of consensus"
repeat_region	12248.. 12291	/note="MERSA repeat: matches 132. .189 of consensus"
repeat_region	12443.. 12595	/note="MERSA repeat: matches 9. .52 of consensus"
misc_feature	complement(13043..13439)	/note="AluJo/FRAM repeat: matches 162. .310 of consensus"
repeat_region	13397.. 13552	/note="match: GSS: Em:B33790"
repeat_region	14382.. 14420	/note="MIR repeat: matches 35. .193 of consensus"
repeat_region	14421.. 14739	/note="MLT1H repeat: matches 514. .547 of consensus"
repeat_region	14740.. 15085	/note="AluJo repeat: matches 1. .310 of consensus"
repeat_region	16534.. 16892	/note="MLT1H repeat: matches 117. .514 of consensus"
repeat_region	16900.. 17163	/note="FHE1B repeat: matches 1. .364 of consensus"
repeat_region	17207.. 17360	/note="L2 repeat: matches 2491. .2747 of consensus"
repeat_region	18198.. 18489	/note="MLT1H repeat: matches 415. .568 of consensus"
repeat_region	18940.. 19047	/note="AluSg repeat: matches 13. .301 of consensus"
repeat_region	19018.. 19116	/note="MERS8C repeat: matches 1. .132 of consensus"
repeat_region	19103.. 19197	/note="3 copies 33 mer 82% conserved"
repeat_region	19785.. 20098	/note="MERS8C repeat: matches 119. .89 of consensus"
repeat_region	20136.. 20382	/note="AluJ repeat: matches 2. .306 of consensus"
repeat_region	21260.. 21460	/note="MIR repeat: matches 8. .261 of consensus"
repeat_region		/note="MIR repeat: matches 2. .219 of consensus"
repeat_region	22331.. 22644	/note="AluY repeat: matches 1. .302 of consensus"
repeat_region	23110.. 23419	/note="AluSg repeat: matches 3. .312 of consensus"
repeat_region	23996.. 24102	/note="LTR33 repeat: matches 415. .521 of consensus"
repeat_region	24208.. 24360	/note="MIR repeat: matches 38. .192 of consensus"
repeat_region	24986.. 25364	/note="LTR16B repeat: matches 1. .449 of consensus"
repeat_region	25661.. 25780	/note="LTR16C repeat: matches 1. .134 of consensus"
repeat_region	25792.. 26153	/note="MERS2B repeat: matches 272. .631 of consensus"
repeat_region	26154.. 26431	/note="AluSg repeat: matches 1. .278 of consensus"
repeat_region	26432.. 26701	/note="MERS2B repeat: matches 2. .272 of consensus"
repeat_region	26702.. 26928	/note="LTR16C repeat: matches 150. .387 of consensus"
repeat_region	27502.. 27831	/note="AluJb repeat: matches 1. .312 of consensus"
repeat_region	27890.. 27960	/note="L1MC4 repeat: matches 7716. .7789 of consensus"
repeat_region	27988.. 28097	/note="L1MC4 repeat: matches 7857. .7977 of consensus"
misc_feature	complement(29592..30009)	/note="match: STS: Em:G21248"
polyA_signal	29986.. 29991	/gene="ba149i18.1"
polyA_site	30009	/gene="ba149i18.1"
Query Match 37.2%; Score 519; DB 9; Length 84122;		
Best Local Similarity 100.0%; Pred. No. 4.2e-81;		
Matches 519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy 877	gacacagacagctgtgagcgtgagctgagctgcaaaagcagagttcttaagaagtacatg 936	
Db 28357	GACACACACAGCTGTGAGCGCTGGATGAGCTGCARAAAGCGAGTTCCTTAAGAAGTACATG 28416	
Qy 937	cacaagtgatgaatgacctgccagctgccctgtctctaccacactgaggtggcctac 996	
Db 28417	CACAAGTGATGAATGACCTGCCAGCTGCCCTGTCTTACCCACTGAGGTGGCTAC 28476	
Qy 997	agcagcgccacatcttcgaccgcatcaagcgaaggaacttcgctggaagcgcagc 1056	
Db 28477	AGCACGCCGACATCTTCGACCGCATCAAGCGCAAGGACTTCCGCTGGAAGGACGCCG 28536	
Qy 1057	gggcccaagagaagctggagattcaagcccaactgcccgtactgcatccctccatg 1116	
Db 28537	GGGCCCCAAGGAGAGCTGGAGATCTACAGCCCACTGCCCGGTACTGCATCCGCTCCATG 28596	
Qy 1117	ctgtccctggagagcacacagctggcgacagcactgctgctacgcgcaacaatgcag 1176	
Db 28597	CTGTCCCTGGAGAGCACACAGCTGGCGGCACAGCAGCTGCTGCTACGGCGCAACATCG 28656	
Qy 1177	ctcatcaccaggggcaagggggcgccagcccaactcatcagcacgagttctcccg 1236	
Db 28657	CTCATCACAGGGGCAAGGGGGCGGCACGCCCACTCATCAGCAGCCGAGTTCCTCCGCG 28716	
Qy 1237	gagctccactacaaggtggacgttctcctgcctggattatctgcaaggggtgactggagcagg 1296	
Db 28717	GAGTCCCACTACAGGTGGAGCTCTGCCCTGGATTATCTGCAAGGGTGTGCTGAGCAGG 28776	
Qy 1297	tatacagagggccggcctcccaacaacggacagaagtgcacagagagccctcgagcag 1356	
Db 28777	TATAACGAGGCGCGCTCTCCCAACAACGACAGAGTGCACAGAGGCCCTCGGACGAG 28836	
Qy 1357	gactacatcaagcaggttcccaagagccaggaatatataa 1395	
Db 28837	GACTACATCAAGCAGTTCCAAAGAGGCCAGGAATATTTAA 28875	

```
RESULT 4
AX351004
LOCUS AX351004 495 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 3 from Patent WO0190179.
ACCESSION AX351004
VERSION AX351004.1 GI:18616373
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Turner,C.A., Hilbun,E., Donoho,G., Friedrich,G.C., Abulin,A.,
Zambrowicz,B. and Sands,A.T.
TITLE Novel human thrombospondin-like proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0190179-A 3 29-NOV-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source 1..495
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 126 a 136 c 139 g 74 t
ORIGIN

Query Match 35.5%; Score 495; DB 6; Length 495;
Best Local Similarity 100.0%; Pred. No. 8.8e-77;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 901 atgagctgcacaaagcaggtctttaaagaagcacatgcacaggtgatgaatgacctgcc 960
Db 1 ATGAGCTGCACAAAGCGAGTCTTAAAGACGTACATGTCACAGGTGATGATGACCTGCC 60

Qy 961 agctgccccctgctctaccaccactgagtgccctacagcagcggcgacatcttcgaccgc 1020
Db 61 AGCTGCCCTGCTCTACCCCACTGAGTGGCTTACAGCAGCGCGGCGACATCTTCGACCGC 120

Qy 1021 atcaagcgaaggactccctcggaagacagcagcggcccaaggagaagactggagatc 1080
Db 121 ATCAAGCGCAAGGACTTCCCGTGAAGACGCCAGCGGGCCCAAGGAGAGCTGGAGATC 180

Qy 1081 tacaagccactgcccgtactgcatccctccatgctgtccctggagagcaccacgctg 1140
Db 181 TACAAGCCCACTGCCGGTACTGTCATCCGCTCATGCTGTCCTCTGGAGAGCACCACGCTG 240

Qy 1141 gcggcacagcactgctgctaaggcgacaacatgcagctcatcacccaggggcaaggggcg 1200
Db 241 CGGCACAGCACTGCTGTACGGCGGACAAACATGCAGCTCATCACAGGGGCAAGGGGCG 300

Qy 1201 ggcagcccaactcatcacagcaggttctccgagagctccactacaaggtgacgtc 1260
Db 301 GGCAGGCCCAACCTCATCAGACCCGAGTCTTCCGGGAGCTCCACTACAAGGTGGAGCTC 360

Qy 1261 ctgcccctggattatctgcaagggtgactggagcaggtataacagagccggcctcccaac 1320
Db 361 CTGCCCTGGATTATCTGCAAGGTGACTGGAGCAGGTATACAGAGCCCGGCTCCCAAC 420

Qy 1321 aacggacagaagtgcacagagagccctcggacagagactacatcaagcagttccaagag 1380
Db 421 AACGGACAGAAGTGCACAGAGAGCCCTCGGACGAGGACTACATCAAGCAGATTCCCAAG 480

Qy 1381 gccagggaattattaa 1395
Db 481 GCCAGGGAATTATAA 495

RESULT 5
BC017997
LOCUS BC017997 1617 bp mRNA linear PRI 06-DEC-2001
DEFINITION Homo sapiens, clone IMAGE:4252124, mRNA, partial cds.
ACCESSION BC017997
```

```
BC017997.1 GI:17389973
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1617)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
COMMENT NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILLNL at: http://image.llnl.gov
Series: IRAL Plate: 35 Row: f Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
Location/Qualifiers
1..1617
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4252124"
/tissue_type="Skin, melanotic melanoma, high MDR."
/clone_lib="NIH_MGC_62"
/lab_host="DH10B"
/notes="vector: pDNR-LIB"
<1..457
/codon_start=2
/product="Unknown (protein for IMAGE:4252124)"
/protein_id="AAH17997.1"
/db_xref="GI:17389974"
/translation="KVMNLSLPCSPYPTVEYVSTADIFRIKRPKDFRWKDSGPKK
LEIYKPTARYCIKRMUSLSTTTAAQHCCYGDNNQLITRGKAGTPLNLISTFSAELH
YKVDVLPWIIICKDWSRYNEARPPNNQKCTESPSDEDIYKQFEAREY"

BASE COUNT 506 a 348 c 391 g 372 t
ORIGIN

Query Match 32.8%; Score 457; DB 9; Length 1617;
Best Local Similarity 100.0%; Pred. No. 3.5e-70;
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 939 caagtgatgaatgacctgccagctgccctgctctctacccactgagtgacctacag 998
Db 1 CAAGGTGATGAATGACCTGCCAGCTGCCCTGCTCTCTACCCCACTGAGGTGGCTACAG 60

Qy 999 cagggcgacatcttcgaccgcatcaagcgcaaggactccgctggaaggagcgccagcg 1058
Db 61 CACGGCGACATCTTCGACCCGCATCAAGCGCAAGGACTTCCGCTGGAAGGACGCGCGG 120

Qy 1059 gcccaaggagaagctggagatctacaagcccaactgcccggctactgcatccgctccatgct 1118
Db 121 GCCCAAGGAGAAGCTGGAGATCTACAAGCCCACTGCCCGGTACTGCATCCGCTCCATGCT 180

Qy 1119 gtccctggagagcaccacgctggtggcgacagcactgctgtctacgagcgacacatgacgt 1178
Db 181 GTCCCTGGAGAGCACCAGCTGTGGCGGCACAGCACTGTGCTACGCGGCAACATGACGCT 240
```





```

/rpt_family="5S"
complement(6715..6815)
/note="low quality data"
complement(6914..7458)
/rpt_family="LINE2"
7523..7707
/rpt_family="L1MB6"
7708..8007
/rpt_family="AluSx"
8233..8305
/rpt_family="(TA)n"
9186..9230
/rpt_family="(CA)n"
complement(join(9248..9470,10060..10240,16529..16722,
25330..25484))
/note="Intron-exon boundaries defined in relation to an
EST contigs that includes AA397580, A1138314, AA399545,
AA86184, all from testis. The closest match in BLASTX is
to E. coli pyruvate-5-carboxylate reductase."
/codon_start=1
/product="unknown"
/protein_id="AAD09625.1"
/db_xref="GI:4186186"
/translation="MDMLQDLESLOFEYGPEDRWLYLQGRSLMIEACAHATFF
CKLYNLRASLNKNQSRHLSIGLSNATPEEFKVLGGHGLGQLAGTLQLGPPI
AESLRISTRPRTIGELQKIGKCFYHNADVSWADVFCLCLPSQLPNCVBIYTSL
EKASIVSVAAIPLRLKLLNHTNLRPOYQYDEDSVSWGANKGVIAALDPTIL
QATCPYPAGHGSYFWHSSEYFPVGPLLSY"
complement(9631..9708)
/rpt_family="LINE2"
complement(10461..10762)
/rpt_family="AluSx"
complement(10804..11077)
/rpt_family="AluJb"
complement(11156..11306)
/rpt_family="MIR"
complement(11493..11615)
/rpt_family="FLAM_C"
11800..12161
/rpt_family="THE1C"
complement(12406..12695)
/rpt_family="AluSx"
12746..12920
/rpt_family="BC200"
12921..13217
/rpt_family="AluSx"
13236..13751
/rpt_family="L1ME1"
13783..14084
/rpt_family="AluSx"
14085..14403
/rpt_family="L1ME2"
14405..14717
/rpt_family="AluSx"
14795..14857
complement(14884..14905)
/rpt_family="AT_rich"
complement(15106..15393)
/rpt_family="AluSx"
complement(15606..15818)
/rpt_family="L1ME2"
complement(15835..16131)
/rpt_family="AluSc"
complement(16775..17073)
/rpt_family="AluY"
17491..17777
/rpt_family="AluJo"
17778..17804
/rpt_family="(CA)n"
complement(17815..17882)
/rpt_family="L1MB2"
17883..18175

```

```

repeat_region
18177..18251
/rpt_family="(TAGA)n"
complement(18244..18306)
/rpt_family="L1MB1"
repeat_region
complement(18310..18609)
/rpt_family="AluSx"
repeat_region
complement(18624..18771)
/rpt_family="L1MB3"
repeat_region
complement(18765..18847)
/rpt_family="L1"
18858..19501
/rpt_family="SVA"
complement(19503..20417)
/rpt_family="L1"
complement(20104..20625)
/rpt_family="L143_Send"
repeat_region
complement(20724..21029)
/rpt_family="AluJb"
repeat_region
complement(21030..21122)
/rpt_family="(TAAA)n"
complement(21125..21423)
/rpt_family="AluJo"
complement(21486..21662)
/rpt_family="L191_Send"
repeat_region
complement(21664..21966)
/rpt_family="AluSx"
repeat_region
complement(22042..22342)
/rpt_family="AluY"
complement(22350..22937)
/rpt_family="L143_Send"
repeat_region
complement(22975..23090)
/rpt_family="FLAM_C"
repeat_region
complement(23164..23287)
/rpt_family="L143_Send"
repeat_region
complement(26429..26704)
/rpt_family="L1MB5"
26766..27080
/rpt_family="AluSx"
complement(28232..28432)
/rpt_family="MER20"
28528..28689
/rpt_family="AluJb"
28692..28984
/rpt_family="AluSx"
complement(29016..29108)
/rpt_family="MIR"
complement(join(30211..30231,30965..31069,31601..31690,
32276..32362,36437..36526,36857..36898,37854..37988,
38439..38514,40403..40476,43661..43688,45155..45257,
45714..45747,46844..46936,51089..51145,51888..51952,
53782..53928,55894..55996,56605..56697))
/note="Intron-exon boundaries defined by an EST contig
that includes W33552, W80307, AA780270, AA278570, R53310,
and AA437679. There is a short form of this coding
sequence that terminates at position 188433, based on EST
AA780270. The closest similarity by BLASTX is to a C.
elegans hypothetical protein."

```

```

Query Match      19.3%; Score 268.8; DB 9; Length 231464;
Best Local Similarity 70.3%; Pred. No. 1.7e-37;
Matches 360; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

Qy  893 gacagctgagcgtgagctgacagctgcaaaagcaggtcttaagaagatcatcaccaag 942
      |||||
Db  78702 GACAGCTGTGAGAGTGGCTGAACGTGCAAGAGCGGCTTCTTAATCAAGTATCTGAGCCAG 78643
      |||||

Qy  943 gtgatgaatgacctgccagctgccctgtctctaccacctgaggtggtgacctagcagc 1002
      |||||
Db  78642 ATGCTGGGGACCTGCCACGCTGCCGCTGCTTACCTACCATGGAGGCCATGACAGCCCT 78583
      |||||

Qy  1003 gccgacatcttcgaccgcacatcaagcgcaaggaactccgctggaagcgccagcgcc 1062
      |||||

```





```

repeat_region 12529..13073
/note="L1MA5A repeat: matches 3652..4242 of consensus"
13076..13330
/note="L1MA5A repeat: matches 4230..4487 of consensus"
13331..13638
/note="L1MA5A repeat: matches 1..307 of consensus"
13639..15444
/note="L1MA5A repeat: matches 4487..6295 of consensus"
15775..15814
/note="10 copies 4 mer tgtg 95% conserved"
complement(16370..16537)
/note="match: GSS: Em:AQ044245"
17517..17817
/note="L1MA5A repeat: matches 82..385 of consensus"
18048..18267
/note="L1MA5A repeat: matches 12..247 of consensus"
18343..18706
/note="match: GSS: Em:AQ099012"
18517..18592
/note="L1MA5A repeat: matches 2370..2445 of consensus"
19038..19141
/note="L1MA5A repeat: matches 2593..2709 of consensus"
19323..19810
/note="L1MA5A repeat: matches 31..547 of consensus"
20121..20430
/note="L1MA5A repeat: matches 1..310 of consensus"
20475..20589
/note="L1MA5A repeat: matches 2634..2747 of consensus"
20991..21056
/note="L1MA5A repeat: matches 71..121 of consensus"
21203..21290
/note="L1MA5A repeat: matches 419..502 of consensus"
21404..21607
/note="L1MA5A repeat: matches 5265..5476 of consensus"
22216..22765
/note="L1MA5A repeat: matches 104..657 of consensus"
22854..23840
/note="L1MA5A repeat: matches 653..1691 of consensus"
23841..24130
/note="L1MA5A repeat: matches 1..308 of consensus"
24131..24246
/note="L1MA5A repeat: matches 1691..1794 of consensus"
24461..25129
/note="L1MA5A repeat: matches 4023..4710 of consensus"
25135..25562
/note="L1MA5A repeat: matches 1..434 of consensus"
25563..26665
/note="L1MA5A repeat: matches 5024..6151 of consensus"
26661..28260
/note="L1MA5A repeat: matches 4697..6294 of consensus"
28280..28413
/note="L1MA5A repeat: matches 248..357 of consensus"
28416..28624
/note="L1MA5A repeat: matches 1794..2012 of consensus"
28628..28911
/note="L1MA5A repeat: matches 1..286 of consensus"
28912..28939
/note="14 copies 2 mer ca 89% conserved"
28955..29226
/note="L1MA5A repeat: matches 1..252 of consensus"
29228..29703
/note="L1MA5A repeat: matches 2090..2208 of consensus"
29922..30344
/note="L1MA5A repeat: matches 111..516 of consensus"
30917..30976
/note="30 copies 2 mer tg 91% conserved"
31509..31556
/note="L1MA5A repeat: matches 33..80 of consensus"
31733..32100
/note="L1MA5A repeat: matches 3..387 of consensus"
32790..32873
/note="L1MA5A repeat: matches 2359..2451 of consensus"
32897..32955

```

```

repeat_region /note="Charlie5 repeat: matches 1..59 of consensus"
32951..33162
/note="Charlie5 repeat: matches 2337..2563 of consensus"
34609..34923
/note="MLT1A1 repeat: matches 1..316 of consensus"
34924..35002
/note="L1MB1 repeat: matches 6076..6151 of consensus"
35114..35281
/note="42 copies 4 mer tctt 78% conserved"
35120..35248
Query Match 13.1%; Score 183.2; DB 9; Length 107057;
Best Local Similarity 98.4%; Pred. No. 1.5e-22;
Matches 185; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 132 gctgcagaataacctcaacgttggaagtgacacccatcagaaacagctttctctc 191
Db 106870 GTTTAGATAACCTCAACGTGGGAAGTGACACCATCAGAAACAGCTTTCTCTC 106929
QY 192 caaagaagcaccacgaagggagcatctggacacccaggtgacacccacccctccagacc 251
Db 106930 CAAAGAAGCACCACGAAGGAGCATCTGGACCAACCAAGGCTGCCACCAACCTTCCCGAGACC 106989
QY 252 gcatcccaagacagacagggcgaccccttcattgcaagagattcccccagatcccttct 311
Db 106990 GCGATTCGACAAAGACAGGGGGGACCCCTTCATTGCAAGAGATTCCCGAGATCCTTTCT 107049
QY 312 ccttgatc 319
Db 107050 CCTTGATC 107057
RESULT 10
HS107712S/c 523 bp DNA linear STS 24-AUG-1999
LOCUS H.sapiens STS from genomic clone 107712, sequence tagged site.
DEFINITION AL110084
ACCESSION AL110084.1 GI:5777434
VERSION STS.
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 523)
AUTHORS Deloukas,P., O'Neill,L., Holden,J., Mistry,D., Huckle,E., Taylor,R.
and Hunt,S.
TITLE Direct Submission
JOURNAL Submitted (24-AUG-1999) E-mail contact: humquery@sanger.ac.uk
COMMENT Marker stdJ107712Sp6 (Primer A : GGAAATCTCTTGGCAATGAAGG; Primer B :
ACCTCAAGTGGGAAGTGAC; amplicon size : 156 bp) is from sequence
generated from the SP6 end of PAC 107712. 107712 is part of the
bacterial clone contigs constructed by the Chromosome 20 Mapping
Group. (http://www.sanger.ac.uk/HGP/Chr20/) 107712 is from the
library constructed at the Roswell Park Cancer Institute by the
group of Pieter de Jong.
For further details see http://bacpac.med.buffalo.edu/.
FEATURES
source
1..523
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="107712"
BASE COUNT 150 a 102 c 116 g 155 t
ORIGIN
Query Match 11.8%; Score 164.2; DB 11; Length 523;
Best Local Similarity 98.2%; Pred. No. 4.3e-19;
Matches 166; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 132 gctgcagaataacctcaacgttggaagtgacacccatcagaaacagctttctctc 191
Db 169 GTTTAGATAACCTCAACGTGGGAAGTGACACCATCAGAAACAGCTTTCTCTC 110

```





Kato, K.  
 Direct Submission  
 Submitted (03-DEC-1993) Ko Kato, Osaka University, Department of  
 Biotechnology; 2-1 Yamada-oka, Suita, Osaka 565, Japan  
 (E-mail: tyomodo@dbj.nig.ac.jp, Tel: 06-877-5111 (ex. 4372),  
 Fax: 06-876-2250)  
 2 (sites)  
 Okada, H., Negoro, S., Kimura, H. and Nakamura, S.  
 Evolutionary adaptation of plasmid-encoded enzymes for degrading  
 nylon oligomers  
 JOURNAL  
 Nature 306 (5939), 203-206 (1983)  
 MEDLINE  
 84068129  
 3 (sites)  
 Tsuchiya, K., Fukuyama, S., Kanzaki, N., Kanagawa, K., Negoro, S. and  
 Okada, H.  
 High homology between 6-aminohexanoate-cyclic-dimer hydrolases of  
 Flavobacterium and Pseudomonas strains  
 J. Bacteriol. 171 (6), 3187-3191 (1989)  
 89255082  
 4 (sites)  
 Negoro, S., Kakudo, S., Urabe, I. and Okada, H.  
 A new nylon oligomer degradation gene (nylC) on plasmid pOAD2 from  
 a Flavobacterium sp  
 J. Bacteriol. 174 (24), 7948-7953 (1992)  
 93094123  
 5 (sites)  
 Kato, K., Ohtsuki, K., Koda, Y., Maekawa, T., Yomo, T., Negoro, S. and  
 Urabe, I.  
 A plasmid encoding enzymes for nylon oligomer degradation:  
 nucleotide sequence and analysis of pOAD2  
 Microbiology 141 (Pt 10), 2585-2590 (1995)  
 96036217  
 6 (bases 1 to 45519)  
 Kato, K., Ohtsuki, K., Koda, Y., Maekawa, T., Yomo, T., Negoro, S. and  
 Urabe, I.  
 Structural analysis of nylon oligomer degradative plasmid poad2:  
 whole nucleotide sequence of poad2  
 Unpublished (1993)  
 7 (sites)  
 Kato, K., Ohtsuki, K., Mitsuda, H., Yomo, T., Negoro, S. and Urabe, I.  
 Insertion sequence is6100 on nylon oligomer degradative plasmid  
 poad2  
 Unpublished (1994)  
 Location/Qualifiers  
 1..45519  
 /organism="Flavobacterium sp."  
 /plasmid="plasmid pOAD2"  
 /strain="KI723T1"  
 /db\_xref="taxon:239"  
 1..859  
 /standard\_name="IS6100"  
 /note="RS-IA repeat sequence; The right part of IS6100 in  
 the RS-IA region; insertion sequence"  
 /citation=[6]  
 complement(2811..3989)  
 /gene="nylB"  
 complement(2811..3989)  
 /gene="nylB"  
 /standard\_name="EII"  
 /EC\_number="3.5.1.46"  
 /note="RS-IA repeated sequence; Sequence homologous to  
 RS-IB region; The EII enzyme is 100 times more active  
 toward the substrate than the EII' enzyme."  
 /citation=[2]  
 /codon\_start=1  
 /transl\_table=11  
 /label="RS-IA"  
 /product="6-aminohexanoate-dimer hydrolase"  
 /protein\_id="BAA05087.1"  
 /db\_xref="GI:488340"  
 /translation="MNAKSTGQHPARYPGAAGEFTLDSWQEPHNWFAHLGEMV  
 TAAVSRDPATPAEPVRLDLEQRLTECTDAFLVLRGSEVLAETVYRAGF  
 APDRHLLMSVSKSLCGTVGALIDEGRIDPAQPVTEYVPELAGSVLDGSLVQLDLM

QISIDYNEVDVPASEVQTHDRSAGWTRRGDPADTYEELTTLRGDGGTGEYCSEA  
 NTVLAWIVERTVGLRYVEALSTYLWAKLDADRDATITVDTQGTGFRANGVYSTARDL  
 ARVGRMLDGGVAGPGRVVSQGWVSVLAGSREADMTDEGTFSAFPGSGSYTROMWCTG  
 NERGNVSGIGIHGQNLWLDPRDTSVIVKLSSWPPDTRHWHGLQSGILLDVSRAILDV  
 "

complement(3997..4000)  
 /note="Shine-Dalgarno sequence"  
 complement(4602..5669)  
 /gene="nylC"  
 complement(4602..5669)  
 /gene="nylC"  
 /standard\_name="EIII"  
 /citation=[4]  
 /codon\_start=1  
 /transl\_table=11  
 /label="EIII"  
 /product="endo-type 6-aminohexanoate oligomer hydrolase"  
 /protein\_id="BAA05088.1"  
 /db\_xref="GI:488341"  
 /translation="MNTTPVHALTDIGGIAVDPAAPRLAGPPVGGGNDADFLLAPVR  
 STGREMLRDFDPFGVSGIAHYEEGPTGATVIHYPAGARTAVDARGVAGVSGGYDFNH  
 AICLAGGAGYGLEAGAGVSDALLERLHRTQFAELQVSSAVIYDFARSATVYDPKA  
 LGRALRPAVPGEPQGRAGAGMSACKVDMDRTETITGQGAFFRLGDRVILAVVVP  
 NPVGIVDRACTVVRGNYDAOTGVRRHPVDYQEAFAEOVPVPTTEAGNTTISATVTV  
 RMPVELNQFAKQVHSSMRGIPQFHTDMDGDTLEAVTTDEIDLPTTPGSSSRGRLSVN  
 ATALGAIASEVMWDVLEAGK"  
 complement(5674..5678)  
 /note="Shine-Dalgarno sequence"  
 14305..15190  
 /standard\_name="IS6100"  
 /note="RS-IB repeat sequence; insertion sequence"  
 /citation=[6]  
 /label="RS-IB"  
 15191..15604  
 /note="RS-IB region; direct repeat from right end of  
 insertion sequence 6100 (IS6100)"  
 /cpt\_type=direct  
 15858..15861  
 /note="Shine-Dalgarno sequence"  
 15869..17047  
 /gene="nylB"  
 15869..17047  
 /gene="nylB"  
 /standard\_name="EII'"  
 /EC\_number="3.5.1.46"  
 /note="RS-IB repeated sequence; Sequence homologous to  
 RS-IA region"  
 /citation=[2]  
 /codon\_start=1  
 /transl\_table=11  
 /label="RS-IB"  
 /product="6-aminohexanoate-dimer hydrolase"  
 /protein\_id="BAA05089.1"  
 /db\_xref="GI:488342"  
 /translation="MNTPTTGSHPARYPDAAAGEPTLDSWQEPHNWFAHLGEMV  
 SRAVSRPVNAPCHALRIGATAAOLPDLEQLEOTYDAFLVLRGTEVVAEYVYRAGF  
 APDRHLLMSVSKSLCGTVGALVDEGRIDPAQPVTEYVPELAGSVLDGSLVQLDLM  
 QISIDYNEVDVPASEVQTHDRSAGWTRRGDPADTYEELTTLRGDGGTGEYCSEA  
 NTVLAWIVERTVGLRYVEALSTYLWAKLDADRDATITVDTQGTGFRANGVYSTARDL  
 ARVGRMLDGGVAGPGRVVSQGWVSVLAGSREADMTDEGTFSAFPGSGSYTROMWCTG  
 NERGNVSGIGIHGQNLWLDPRDTSVIVKLSSWPPDTRHWHGLQSGILLDVSRAILDV

17849..18728  
 /standard\_name="IS6100"  
 /note="RS-IC repeat sequence; insertion sequence"  
 /citation=[6]  
 /label="RS-IC"  
 complement(29155..30034)  
 /standard\_name="IS6100"  
 /note="RS-ID repeat sequence; insertion sequence"  
 /citation=[6]  
 /label="RS-ID"

misc\_feature  
 misc\_feature  
 misc\_feature





**Notes:**

Corney, Norwich, Norfolk NR9 7QH, UK

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics

Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide web.

(URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/))

CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

```
/gene="SCC88.02"
3580. 5310
/notes="SCC88.02"
similar to TR: CAB51964 (EMBL:AL109661) Streptomyces
coelicolor hypothetical 100.7 kD protein SC6E10.10, 943
aa; fasta scores: opt: 522 z-score: 556.2 E(): 1.4e-23;
31.8% identity in 471 aa overlap. Contains possible
N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAB75372.1"
/db_xref="GI:6983733"
/translation="MTFSRLRLTRSLTCTALAVAGMLVTTTAVAADKPSGIE
VPTKSLPLDRAPQKPKAPMRRAATAPRFDVGDGFSMLRLQDSTQVGLT
TADGFTYITNVDDYERKQDVITPGLDNGGPEIILSSSTGLSLFQSGWADSTGY
ATWSGTQYINQYKVPAGDVTGDRPDLARTYGGDLYLYQGTGSDSPRGRVKG
GWEMFDQVGMNGDGIQDVVARTLGLYLYFAGKSAATLATRLDLGAGWYIN
QIIAADQVGMNGDGLGRTSGDLYLYRADTGNFEAKQOAGTGPVPEFFAGSGY
PHFCASLVGMVTRGTYLYWVKNQDGLTGREOISDTGWSGARVAYASLDLGYAD
LLEYDGLVNYAATADESLATGKYNLFLPGDLNDCKGDLILDARTSGTLYFRG
KHGASLSPLKVGCGWNAVDKLVGADGTFDGRDITVARTPAGKLYLRYGTGVSNSP
FAGKVDLGSQWQYNKLASPDIDGGRADLLAVNSGGDLYRTSTYDGRFARVILG
TGMWYRNLY"
5327. 5362
complement(5387..8215)
/notes="SCC88.03c"
/gene="SCC88.03c"
complement(5387..8215)
/notes="SCC88.03c"
possible transferase, len: 942 aa;
N-terminal region similar to hypothetical proteins, e. g.
TR: 069853 (EMBL:AL023702) Streptomyces coelicolor
hypothetical 66.0 kD protein SC1C3.11, 586 aa; fasta
scores: opt: 1214 z-score: 1395.6 E(): 0; 40.7% identity
in 573 aa overlap and C-terminal region similar to
TR: 069854 (EMBL:AL023702) Streptomyces coelicolor putative
transferase SC1C3.12, 697 aa; fasta scores: opt: 658
z-score: 754.0 E(): 0; 30.1% identity in 574 aa overlap.
Contains Pfam match at N-terminal region to entry PF00534
Glycosyl_transf_1, Glycosyl transferases group 1"
/codon_start=1
/transl_table=11
/product="putative transferase"
/protein_id="CAB75373.1"
/db_xref="GI:6983734"
/translation="MKITFLTWGDMGTRQAVVYTOATHLAPRHDIEVLSVFKTRQ
PFFSVDSVSVRYLVDRGKTERPVRESDLTAEDCRHLAGLPSDLISPKWEATFDALS
DVEQMRALRTIDTVLITSPALMSAVADLAAPSRTITQOEHRPSQLRGTEGPELLLR
LPAIDALVLTERTKQWLEESLGRKAPRLAIPNAIPEGFRPSRLSGTKTIEMPRRLV
FDKQVDHAIQAFKALPDHGRURKIFGDPQMSRLRLNLIQGLHSDSVELLGFSQHM
TEEMARASLITLPSQDGEAFPLVLEAFAGVPVAVYDIVTGPAEIIHGEDGLLVPP
NDVESLAERISRLMGDEALLRSYGEKAHEGSTREAAADVIVKQWEELEFTEVSRDDPR
RMAERADRIAHVAGGAGRFHAAVAADRTAPSSGDORAREWIGASDRSLVRAGRL
SFVRDLOGSEIVQNFETVVEALESIGIPVLLDRDDNPRRLVADAEOTVRKA
LAGATEGRAVTAELKPRTHAPGVLLAERLEAVEGAVGLRFPVPTSTRLRGPAY
GCDIEFVRQVEEGEGDGVAPRPSAVGPKLPLETDARTVRKDRFYPELPLTRK
LVSDFITFPVDVYTWDDSDPRWQERRARRAALGLEAESDGEAAERFNRDELRLSL
RSLHAFMWIRKIYLVITDDQTPWLNTEHEGIEVYSHRDIPTDQDCLPTFNHSIESQ
LHHIDGLSQFQILVNDVFIQRPVGAORFFLPNGASRFVPSPTVPVGEPTDEDEGYF
AAAKNNRALLERFGVTVANSFVHAPHLRBSVLSEIEEDFESVARTAAATPMRGWQD
ISMWLSLHHHYGLFGASVPSIRCAIYIDVGTYSRHPELTLLAMRHGHDVFCLESQD
AEVPEHQARIVAEAFRLAFYFPVKSPYER"
complement(7124..7570)
/notes="SCC88.03c"
/notes="Pfam match to entry PF00534 Glycosyl_transf_1,
Glycosyl transferases group 1, score 99.70, E-value
2.1e-27"
complement(8222..9211)
/gene="SCC88.04c"
complement(8222..9211)
/gene="SCC88.04c"
/notes="SCC88.04c, unknown, len: 329 aa"
```

```

gene
Query Match 3.8%; Score 52.4; DB 1; Length 42816;
Best Local Similarity 45.3%; Pred. No. 10;
Matches 191; Conservative 0; Mismatches 231; Indels 0; Gaps 0;
QY 942 ggtgatgaatgacctgcccagctgcccctctctaccctgaggtgacctacagcac 1001
Db 32480 GGTGATGAGGCCAAGACCGCGCGCTAGTCGCCCTCCAACCCCGACTACGACCC 32421
QY 1002 ggcgcacatcttcgaccgcatcaagcgcaagacctctcgtggaagacgcccagcgcc 1061
Db 32420 CAACGCCCTGGTTCGGCGGATCTCCGCCAAGGACTACAAGAAGCTACCCGCGAGGACTC 32361
QY 1062 caaggaagaagtggagatctacaagccactgcccgtgactgcatcgctccatgctgtc 1121
Db 32360 CAACTACCCGCTGCTCAACCGCGCCACCCAGGCGCCAGCGCGCTCCATCTTCAA 32301
QY 1122 cctggagagacacacgctggcggcacagacactgctgtacggcgacaacatgagctcat 1181
Db 32300 GGTGCTCTCTCGCGCGAGCGGTCGAGGCGCGCTAGACTTCGACGGCAACTACGACTG 32241
QY 1182 caccaggggcaagggggcgccgacgcccacacctcatcagcaccgagttctcccgcgagct 1241
Db 32240 CTCGAGCGCCTACAACGTCGGCGCCAGGCTTCAAGAAGCTTCGAGTCGGCCACAGAGGG 32181
QY 1242 ccactacaaggtggagctctcctgctggtattatctgcaaggggtgactggagcaggtataa 1301
Db 32180 CATGATCACCATCGCGCGCGCTCGAGCTCTCTCTGGGACACACCGCTCTTCTACCGCTGTC 32121
QY 1302 cgaggcccgccctcccaacacgacagaaagtgcacagagagagccctcgagcagaggacta 1361
Db 32120 CCACGAGGTGGAGAGGACGCGGGGATGAACCCCAAGAACCCCGCGACTACTTCTA 32061
QY 1362 ca 1363
Db 32060 CA 32059
```

Search completed: April 25, 2002, 12:13:26  
Job time: 7535 sec

misc\_feature

gene

CDS





```
351 TATCAATGGCGAGRWTCCTCAATATCCAGGTCCACCATAGAGTGTGTCGAGC 400
134 lyProAspSerGluAlaAspLysAspGlnHisProGluAsnLysProSer 150
401 GTCTCTGACTCTGAAGCAGATAAAGATCAGCATCCGAGGAGAAATGAAGCCAGC 450
151 TrpSerValProSerProAspTrpArgAlaTrpTrpGlnArgSerLeuSe 167
451 TGGTCAGTCCCATCCCCAGCATGGCGGGCCCTGGTGGCAGAGTCCCTGTC 500
167 rLeuAlaArgAlaAsnSerGlyAspGlnAspTyr***TyrAspSerThr 184
501 CTTGCCAGGGCAACACAGCGGGACCACTACRAGTACGACAGTACCT 550
184 erAspAspSerAsnPhelLeuAsnProProArgGlyTrpAspHisThrAla 200
551 CAGACGACAGCAACTTCTCAACCCCGCCAGGGGTGGGACCATACAGCC 600
201 ProGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerTh 217
601 CCAGGCCACCGGACTTTTGAACCAACAGATCAGCCAGATATGATTCAC 650
217 rAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerValThrCysG 234
651 AGATGGCGAGGGTGACTGGAGTCTCTGGTCTGTCTGCAGCGTCACCTGCG 700
234 lyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAla 250
701 GGAAGGCCAACCAAGAACCGGACCCGGTCTTGGGTACGCTGTCAGTCA 750
251 ThrGluSerArgThrCysAspArgProAsnCysProGlyIleGluAspTh 267
751 ACAGAATCGAGGACCTGTGACCGTCCAAACTGCCAGGAATTTGAAGACAC 800
267 rPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerGluGlu 284
801 TTTTAGGACAGCTGCCACCGAAGTGAAGTCTGCTTCGGGAACGAGGAGT 850
284 heAsnAlaThrLysLeuPheGluValAspThrAspSerCysGluArgTrp 300
851 TTAATGCCACCAACTGTTTGAAGTTGACACAGACAGCTGTGAGCGCTGG 900
301 MetSerCysLysSerGluPheLeuLysLysTyrMetHisLysValMetAs 317
901 ATGAGCTGCAAAAGCGAGTCTTAAGAAGATACATGCACAAGGTGATGAA 950
317 nAspLeuProSerCysProCysSerTyrProThrGluValAlaTyrSerT 334
951 TGACCTGCCAGCTGCCCTCTCTTACCCACTGAGGTGGCTACAGCA 1000
334 hrAlaAspIlePheAspArgIleLysArgLysAspPheArgTrpLysAsp 350
1001 CGGCGGACATCTTCGACCGCATCAAGCGCAAGGACTTCCGCTGGAAGGAC 1050
351 AlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAlaArgTy 367
1051 GCCAGCGGGGCCCCAAGGAGAGCTGGAGATCTACAGCCCACTGCCCGGTA 1100
367 rCysIleArgSerMetLeuSerLeuGluSerThrThrLeuAlaAlaGlnH 384
1101 CTGCATCCGCTCCATGCTGTCCCTCGAGAGCACCCAGCTTCCGCTGGAAGC 1150
384 isCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLysGlyAla 400
1151 ACTGCTGTACGGCGCAACATCAGCTCATCACAGGGGGGAAGGGGGGG 1200
401 GlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuHisTyrLy 417
1201 GGCACGCCCAACCTCATCAGCACCGAGTCTCCGCGGAGCTCCACTACAA 1250
417 sValAspValLeuProTrpIleLeuCysLysGlyAspTrpSerArgTyrA 434
1251 GGTGGAGCTCTGCCCTGGATTATCTGCAAGGGTGACTGGAGCAGGTATA 1300
```

```
434 snGluAlaArgProProAsnAsnGlyGlnLysCysThrGluSerProSer 450
1301 ACAGAGCGCGGCTCCCAACACGAGAGAAGTGCACAGAGAGCCCTCG 1350
451 AspGluAspTyrIleLysGlnPheGlnGluAlaArgGluTyr 464
1351 GAGGAGACTACATCAAGCAGTTCACAGAGGCCAGGGAATAT 1392
seq_name: gb_pat:AX351006
seq_documentation_block:
LOCUS AX351006 936 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 5 from Patent WO0190179.
ACCESSION AX351006
VERSION AX351006.1 GI:18616374
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Turner,C.A., Hilbun,E., Donoho,G., Friedrich,G.C., Abuin,A.,
Zambrowicz,B. and Sands,A.T.
TITLE Novel human thrombospondin-like proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0190179-A 5 29-NOV-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Location/Qualifiers
source
1..936
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 231 a 283 c 248 g 171 t 3 others
ORIGIN
```

```
alignment_scores:
Quality: 1567.00 Length: 297
Ratio: 5.366 Gaps: 0
Percent Similarity: 98.316 Percent Identity: 97.980
alignment_block:
US-09-863-824-2 x AX351006 ..
Align seg 1/1 to: AX351006 from: 1 to: 936
1 MetValArgLeuAlaAlaGluLeuLeuLeuLeuLeuGlyLeuLeuLeuLe 17
1 ATGTTGCGCTTGGCGCGGAGCTGCTGCTGCTGGGGCTGCTGCTGCT 50
17 uThrLeuHisIleThrValLeuArgGlySerGlyAlaAlaAspGlyProA 34
51 CACGCTGCACATCACCGTGTGCGGGGCTCGGAGCCGCCGACGGGCCG 100
34 spAlaAlaAlaGlyAsnAlaSerGlnAlaGlnLeuGlnAsnAsnLeuAsn 50
101 ACGCGCGCGCGGCAACCGCACCGCAAGCCAGCTGCAGATAACCTCAAC 150
51 ValGlySerAspThrThrSerGluThrSerPheSerLeuSerLysGluAl 67
151 GTGGGAAGTGACACCATCAGAAACCCAGCTTTCTCTCTCCAAAGAGC 200
67 aProArgGluHisLeuAspHisGlnAlaAlaHisGlnProPheProArgp 84
201 ACCAAGGAGCATCTGGACCACCGCTGCACCAACCCCTCCCCAGAC 250
84 roArgPheArgGlnGluThrGlyHisProSerLeuGlnArgAspPhePro 100
251 CGCGATTCCGACACAGAGACGGGGCACCTTCATTGCAAAAGAGATTCCCC 300
101 ArgSerPheLeuLeuAspLeuProAsnPheProAspLeuSerLysAlaAs 117
301 AGATCCTTTTCTCTTGATCTACCAACACTTTCAGATCTTTTCCAAAGCTGA 350
```



```
117 pileAsnGlyGln***ProAsnIleGlnValThrIleGluValValAspG 134
|||||
351 TATCAATGGCGAGRTWCACAAATATCCAGGTACCATAGAGGTGCTCGAGC 400
|||||
134 lyProAspSerGluAlaAspLysAspGlnHisProGluAsnLysProSer 150
|||||
401 GTCTGAGCTCTGAAGCAGATAAAGATCAGCATCCGGAGAAATAGCCAGC 450
|||||
151 TrpSerValProSerProAspTrpArgAlaTrpTrpGlnArgSerLeuSe 167
|||||
451 TGGTCAGTCCCATCCCCGACTCGGGCCCTGGTGGCAGAGTCCCTGTC 500
|||||
167 rLeuAlaArgAlaAsnSerGlyAspGlnAspTyr***TyrAspSerThrS 184
|||||
501 CTTGGCCAGGCGCAACAGCGGGACAGGACTACRAGTACGACAGTACCT 550
|||||
184 erAspAspSerAsnPhelLeuAsnProProArgGlyTyrAspHisThrAla 200
|||||
551 CAGACGACAGCAACTTCTCAACCCGCCAGGGGTGGGACCATACAGCC 600
|||||
201 ProGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerTh 217
|||||
601 CCAGCCACCGGACTTTTGAACCAACAAAGATCAGCCAGAAATATGATCCAC 650
|||||
217 rAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerValThrCysG 234
|||||
651 AGATGGCGAGGTCGACTGGAGTCTCTGGTCTCTGCAGCGTCACTTCCG 700
|||||
234 lyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAla 250
|||||
701 GGAACGGCAACAGAAAGCGGACCGGCTTGTGGCTACGCGTGCATGCA 750
|||||
251 ThrGluSerArgThrCysAspArgProAsnCysProGlyIleGluAspTh 267
|||||
751 ACAGAATCGAGGACTGTGACCGCTGCCAACTGCCAGGAATTAAGACAC 800
|||||
267 rPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerGluGluP 284
|||||
801 TTTTAGGACAGCTGCCACCGAAGTGAGTCTGCTTGGCGGAAGCGAGGAGT 850
|||||
284 heAsnAlaThrLysLeuPheGluValAspThrAspSerCys 297
|||||
851 TTAATGCCACCAAACTGTTGAAGTGTGCTGCCAGCATGT 891
|||||
```

```
seq_name: gb_pr:AL133463
seq_documentation_block:
LOCUS      AL133463              84122 bp    DNA        linear    PRI 15-MAR-2001
DEFINITION Human DNA sequence from clone RP11-149118 on chromosome 20.
            Contains the 3' end of a novel gene, ESTs, STSS and GSSs, complete
            sequence.
ACCESSION  AL133463
VERSION    AL133463.16   GI:10443352
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 84122)
REFERENCE  1
AUTHORS    Wilton, S.
TITLE      Direct Submission
JOURNAL    Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
COMMENT    On Oct 1, 2000 this sequence version replaced gi:10178396.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
```

numbers given in the feature table with their source databases:  
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human 20  
chromosome 20, constructed by the Sanger Centre Chromosome 20  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr20  
IMPORTANT: This sequence is not the entire insert of clone  
RP11-149118 It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true right end of clone RP4-585114 is at 84023 in this sequence.  
The true left end of clone RP5-107712 is at 100 in this sequence.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e. phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. RP11-149118 is from  
the library RPCI-11.1 constructed by the group of Pieter de Jong.  
For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACE3.6.

```
FEATURES
            Location/Qualifiers
            1..84122
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="20"
               /clone="RP11-149118"
               /clone_lib="RPCI-11.1"
               /join(<1..159,9050..9314,17956..18099,21790..21879,
28358..30067)
               /gene="bAl149118.1"
               /note="match: ESTs: Em:AL191256 Em:R44763 Em:W86073
Em:AI082824 Em:AI886394 Em:R52250 Em:AV708785 Em:W07772
Em:AV660737 Em:BF302850 Em:AW426875 Em:W86257 Em:BF302850
Em:AW873508 Em:BE287313 Em:AW495535 Em:AW435648
Em:AW435639"
               /product="bAl149118.1 (novel protein)"
               /evidence=not_experimental
               1..30067
               /gene="bAl149118.1"
               /join(<1..159,9050..9314,17956..18099,21790..21879,
28358..28875)
               /gene="bAl149118.1"
               /note="continued from dJ107712.1 in Em:AL050320
match: proteins: Tr:O95432 Tr:Q9RTS5"
               /codon_start=1
               /evidence=not_experimental
               /product="bAl149118.1 (novel protein)"
               /protein_id="CAC16127.2"
               /db_xref="GI:13374941"
               /translation="HQAHPFRPRFRQETGHPSLQRDPFRSFLDLNFPDLSKAD
INGQNPQIVTIEVNDPSDEADKQHPENKPSVSPDWMRAWGWSLSLRANSGD
QDYKYDSTSDSNFNLNPPRGWDHATGHFTETKQDQEDYDSTDEGWSLWVCSYTC
GNGNKRTRSCGVACTSRTCDRNPCEIEDFTFAATEVSLLAGSEFNATKLF
VPTDCSRWMSCKSEFLKKYMKVMNDLPSCPCSYPTVEAYSTADIFDRTKRKDFRWK
DASGPKELEIYKPTARYCIRSMISLESTTLAQCCHCCYGDNMQLITRGKACTPNLIS
TEFSALHYKVDVLPWICKGDWSRYNEARPPNNGOKCTSPESDEDIYIKQFQAREY"
               complement(1..81)
               /note="match: STS: Em:HS107712S"
               437..504
               /note="34 copies 2 mer aa 66% conserved"
               445..510
               /note="2 copies 33 mer 93% conserved"
               885..1046
               /note="MIR repeat: matches 35..189 of consensus"
               1202..1494
               /note="AluX repeat: matches 1..291 of consensus"
               2642..2951
               /note="AluJo repeat: matches 1..311 of consensus"
               3460..3507
```

```
misc_feature
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
```

```
repeat_region /note="24 copies 2 mer tg 79% conserved"
misc_feature 4065..4106
repeat_region /note="21 copies 2 mer cc 76% conserved"
misc_feature complement(4442..5302)
repeat_region /note="match: GSS: Em:AQ741486"
misc_feature 4908..5154
repeat_region /note="MIR repeat: matches 34..238 of consensus"
misc_feature 5180..5672
repeat_region /gene="BA1491I18.1"
repeat_region /note="match: GSS: Em:AQ755240"
repeat_region 7941..8067
repeat_region /note="MIR repeat: matches 1..140 of consensus"
repeat_region 8078..8510
repeat_region /note="MLTII repeat: matches 81..519 of consensus"
misc_feature 8538..8624
repeat_region /note="MIR repeat: matches 170..262 of consensus"
misc_feature complement(8621..9044)
misc_feature /note="match: GSS: Em:AQ182554"
repeat_region complement(8758..9044)
repeat_region /note="match: GSS: Em:AQ133736"
repeat_region 9887..10251
repeat_region /note="MLTII repeat: matches 19..371 of consensus"
repeat_region 10536..10673
repeat_region /note="L2 repeat: matches 2563..2708 of consensus"
repeat_region 10826..11052
repeat_region /note="MIR repeat: matches 2..261 of consensus"
repeat_region 11277..11343
repeat_region /note="MER5A repeat: matches 99..166 of consensus"
repeat_region 11344..11517
repeat_region /note="MER5A repeat: matches 1..182 of consensus"
repeat_region 11520..11698
repeat_region /note="MER5B repeat: matches 2..178 of consensus"
repeat_region 11773..11825
repeat_region /note="MER5A repeat: matches 136..187 of consensus"
repeat_region 12156..12212
repeat_region /note="MER5A repeat: matches 132..189 of consensus"
repeat_region 12248..12291
repeat_region /note="MER5A repeat: matches 9..52 of consensus"
repeat_region 12443..12595
repeat_region /note="AluTo/PRAM repeat: matches 162..310 of consensus"
misc_feature complement(13043..13439)
repeat_region /note="match: GSS: Em:B33790"
repeat_region 13397..13552
repeat_region /note="MIR repeat: matches 35..193 of consensus"
repeat_region 14382..14420
repeat_region /note="MLTII repeat: matches 514..547 of consensus"
repeat_region 14421..14739
repeat_region /note="AluTo repeat: matches 1..310 of consensus"
repeat_region 14740..15085
repeat_region /note="MLTII repeat: matches 117..514 of consensus"
repeat_region 16534..16892
repeat_region /note="THE1B repeat: matches 1..364 of consensus"
repeat_region 16900..17163
repeat_region /note="L2 repeat: matches 2491..2747 of consensus"
repeat_region 17207..17360
repeat_region /note="MLTII repeat: matches 415..568 of consensus"
repeat_region 18198..18489
repeat_region /note="AluSq repeat: matches 13..301 of consensus"
repeat_region 18940..19047
repeat_region /note="MER5C repeat: matches 1..132 of consensus"
repeat_region 19018..19116
repeat_region /note="3 copies 33 mer 82% conserved"
repeat_region 19103..19197
repeat_region /note="MER5C repeat: matches 119..89 of consensus"
repeat_region 19785..20098
repeat_region /note="AluY repeat: matches 2..306 of consensus"
repeat_region 20136..20382
repeat_region /note="MIR repeat: matches 8..261 of consensus"
repeat_region 21260..21460
repeat_region /note="MIR repeat: matches 2..219 of consensus"
repeat_region 22331..22644
repeat_region /note="AluY repeat: matches 1..302 of consensus"
repeat_region 23110..23419
```

```
repeat_region /note="AluSq repeat: matches 3..312 of consensus"
repeat_region 23996..24102
repeat_region /note="LTR33 repeat: matches 415..521 of consensus"
repeat_region 24208..24360
repeat_region /note="MIR repeat: matches 38..192 of consensus"
repeat_region 24986..25364
repeat_region /note="LTR16B repeat: matches 1..449 of consensus"
repeat_region 25661..25780
repeat_region /note="LTR16C repeat: matches 1..134 of consensus"
repeat_region 25792..26153
repeat_region /note="MER92B repeat: matches 272..631 of consensus"
repeat_region 26154..26431
repeat_region /note="AluSq repeat: matches 1..278 of consensus"
repeat_region 26432..26701
repeat_region /note="MER92B repeat: matches 2..272 of consensus"
repeat_region 26702..26928
repeat_region /note="LTR16C repeat: matches 150..387 of consensus"
repeat_region 27502..27831
repeat_region /note="AluY repeat: matches 1..312 of consensus"
repeat_region 27890..27960
repeat_region /note="L1MC4 repeat: matches 7716..7789 of consensus"
repeat_region 27988..28097
repeat_region /note="L1MC4 repeat: matches 7857..7977 of consensus"
misc_feature complement(29592..30009)
polyA_signal 29986..29991
polyA_site /gene="BA1491I18.1"
polyA_site /gene="BA1491I18.1"

alignment_scores:
  Quality: 954.00      Length: 204
  Ratio: 5.074        Gaps: 2
  Percent Similarity: 92.157  Percent Identity: 87.745

alignment_block:
US-09-863-824-2 x AL133463

Align seg 1/1 to: AL133463 from: 1 to: 84122

262 ProGlyIleGluAspThrPheArgThrAlaAlaThrGluValSerLeuLe 278
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28276 CCTGGTGC.....ACATGCTCTTGAAGCACCTCCTGTG 28310

278 uAlaGlySerGluGluPheAsnAlaThrLysLeuPheGluVal....AspT 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28311 GGCCGGTTGTACACGCGGTGAGAGGGTTCTCTTTGGCCTTTCCAGACA 28360

294 hrAspSerCysGluArgTrpMetSerCysLysSerGluPheLeuLysLys 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28361 CAGACAGCTGTGCGCTGGATGAGCTGCAAAAGCGAGTCTCTTAAGAAG 28410

311 TyrMetHisLysValMetAsnAspLeuProSerCysProCysSerTyrPr 327
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28411 TACATGCACAAGGTGATGATGACCTGCCAGCTGCCCTCTCTACCC 28460

327 OThrGluValAlaTyrSerThrAlaAspIlePheAspArgIleLysArgL 344
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28461 CACTCAGTGGCGCTACAGACGCGGCATCTTCGACCCGATCAAGCGCA 28510

344 ysAspPheArgTrpLysAspAlaSerGlyProLysGluLysLeuGluLe 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28511 AGGACTTCGCTGGAAGACGCCAGCGGCCCAAGAGAGCTGGAGATC 28560

361 TyrLysProThrAlaArgTyrCysIleArgSerMetLeuSerLeuGluSe 377
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28561 TACAAGCCGCTGCCGGTACTGTCATCGCTCCATGCTGTCTCTGAGAG 28610

377 rThrThrLeuAlaAlaGlnHisCysCysTyrGlyAspAsnMetGlnLeuI 394
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28611 CACCACGCTGGCGGCACAGCACTGCTGTACGGCGCAACATGCAGCTCA 28660

394 lThrArgGlyLysGlyAlaGlyThrProAsnLeuIleSerThrGluPhe 410
```

```

|||||
28661 TCACGAGGGCAAGGGGGGCGGCACGCCCAACCTCATCAGCACCGAGTTC 28710
411 SerAlaGluLeuHisTyrIysValAspValLeuProTyrPheIleCysIy 427
28711 TCCGGGGAGCTCCACTACAAAGGTGGACGTCTCCCTCGATTATCTGCAA 28760
427 sGlyAspTyrSerArgTyrAsnGluAlaArgProProAsnAsnGlyGlnL 444
28761 GGGTCACTGGACAGGTATACGAGCGCGCGCTCCCAACACGACGACA 28810
444 yScyThrGluSerProSerAspGluAspTyrIleLysGlnPheGlnGlu 460
28811 AGTGCACAGAGAGCCCTCGGACGAGGACTACATCAAGCAGTTCCAAGAG 28860
461 AlaArgGluTyr 464
28861 GCCAGGGAATAT 28872

```

seq\_name: gb\_pr:AK056709

seq\_documentation\_block:

LOCUS AK056709 3052 bp mRNA linear PRI 31-OCT-2001  
 DEFINITION Homo sapiens cDNA FLJ32147 fis, clone PLACE5000116.  
 ACCESSION AK056709  
 VERSION AK056709.1 GI:16552189  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens placenta cDNA to mRNA, clone\_lib:PLACES  
 clone:PLACES5000116.

#### ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (sites)  
 Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,  
 Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,  
 Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,  
 Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,  
 Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,  
 WagaTsuna,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,  
 Suzuki,Y., Sugano,S., Nagahara,K., Masuho,Y., Nagai,K. and  
 Isogai,T.

#### TITLE

NEDO human cDNA sequencing project

#### JOURNAL

Unpublished

2 (bases 1 to 3052)

Isogai,T., Otsuki,T. and Sugiyama,T.

Direct Submission

Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,

Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

(E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)

NEDO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; cDNA full insert sequencing:

Research Association for Biotechnology (RAB); cDNA library

construction: Helix Research Institute (HRI) (supported by Japan

Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,

HRI, and Biotechnology Center, National Institute of Technology and

Evaluation; clone selection for full insert sequencing: RAB and

HRI.

Location/Qualifiers

1. .3052

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="PLACE5000116"

/tissue\_type="placenta"

/clone\_lib="PLACES"

/note="cloning vector: pME18SFL3"

622 a 946 c 910 g 574 t

BASE COUNT

ORIGIN

#### alignment\_scores:

Quality: 930.50 Length: 500

Ratio: 2.926 Gaps: 18

Percent Similarity: 63.600 Percent Identity: 41.600

#### alignment\_block:

US-09-863-824-2 x AK056709

Align seg 1/1 to: AK056709 from: 1 to: 3052

```

25 ArgGlySerGlyAlaAlaAspGlyProAspAlaAlaAlaGlyAsnAlase 41
|||||
428 AGAGGTACTCGTTCGGCTGGAGCTCAGAAAGCTCGCGGATGGC... 475
|||||
41 rGlnAlaGlnLeuGlnAsnAsnLeuVal... 51
|||||
476 .....CAACACACCTTGAGTACCCTAACCCCTGATACCCAG 512
|||||
52 .....GlySerAspThr 55
|||||
513 GCTTCAGCCTCCCCAGATCCTAGGCCTCTGAGGGAAGAGGAGGAGGCGCAG 562
|||||
56 ThrSerGluThrSerPheSerLeuSerLysGluAlaProArgGluHisLe 72
|||||
563 ACTGCTCCCCAGACCCACCTCGAGCAGAGCTACACCAACATGGATGTT 612
|||||
72 uAspHisGlnAlaAlaHisGlnProPheProArgProArgPheArgGlnG 89
|||||
613 GGACTGTCTACTGAGCCAGCAGCCCTGACCCAGG.....G 647
|||||
89 lu.ThrGlyHisProSerLeuGlnArgAspPheProArgSerPheLeuLe 105
|||||
648 AATGCCAGCCCTCCAGGACCAGGAGGTACTCC.....TTGCTGCT 691
|||||
105 uAspLeuProAsnPheProAspLeuSerLysAlaAspIleAsnGlyGln* 122
|||||
692 GGAGCTGCAGAGAGCTGCCAGAATTGGTCACCAACCTTGATACCCCTA 741
|||||
122 **ProAsnIleGlnValThrIleGluValAlaSpGlyProAspSerGlu 138
|||||
742 ACCCTGATAAACAGTGCACCAAGTGGTGGAGGAGCCGCCAGGCCGAG 791
|||||
139 AlaAspLysAsp.....GlnHisProGluAsnLysPro..... 149
|||||
792 GTGTGATAGACTCTGTGGCTGAGCCCAATCCCGCCGCCAGGATAC 841
|||||
150 ....SerTrpSerValProSerProAspTrpArgAlaTrpTrp..... 162
|||||
842 CCTTAGCTGG...CTGCCGCCCTC...TGCTCTCTCTCGGGAGACT 885
|||||
163 .....GlnArgSerLeuSerLeuAlaArgAlaAsnSerGly..... 174
|||||
886 ACAAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 935
|||||
175 .....AspGlnAspTyr..... 178
|||||
936 GAGGAAGACGAGGAGTATCTTCAGAGGATATCGAGGTTGAGGATCAAGA 985
|||||
179 .....***TyrAspSerThrS 184
|||||
986 GGACAAAGAGGAGATGAGGAAGAGCAGCGCTCTGGTTCAATGGAACTA 1035
|||||
184 erAspAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAla 200
|||||
1036 CAGAC.....AACTGGGACACAGGCTGG.....CTGGCC 1064
|||||
201 ProGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerTh 217
|||||
1065 CCCGGGAGTGGGTCTTC.....AAGGATCTCTCAGCTAGCAGC...TA 1105
|||||
217 rAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerValThrCysG 234
|||||
1106 TGAGCCTCAGAGGAGTGGAGTCCCTGGTCTCCCTGCAGTGGGAACTGCA 1155
|||||
234 lyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAla 250
|||||
1156 GCACCTGGCAAGCAGCAGAGGAGCTCGGCCCTGTGGCTATGGCTGCAC 1205
|||||

```



```

      :::::   ::   |||   :::::||||
815  ACAAGGAGAGAAACACAGGCCCCCAGGGGAGAGGGGAGGAAAAG 864
      :::::   ::   |||   :::::||||
175  .....AspGlnAspTyr..... 178
      |||:::|||||
865  GAGGAAGACGAGGACTATCTTCAGAGGATATCGAGGCTGAGGATCAAGA 914
      :::::   ::   |||   :::::||||
179  .....AAspGlnAspTyr.....**TyrAspSerThrS 184
      :::::   ::   |||   :::::||||
915  GCACAAAGCAGGAGATGAGGAGACGAGCGCTCTGGTTCAATGGAACATA 964
      :::::   ::   |||   :::::||||
184  erAspAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAla 200
      :::::   ::   |||   :::::||||
965  CAGAC.....AACTGGGACGAGGGCTGG.....CTGGCC 993
      :::::   ::   |||   :::::||||
201  ProGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerTh 217
      |||||:::   ::   |||   :::::||||
994  CCCGGGGATGGGTCTTC.....AAGGATTCTGTCAGCTACGAC...TA 1034
      :::::   ::   |||   :::::||||
217  rAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerValThrCysG 234
      :::::   ::   |||   :::::||||
1035  TGAGCCTCAGAAGGAGTGAGTCCCTGCTCCCTGCAGTGGGAACATGCA 1084
      :::::   ::   |||   :::::||||
234  lyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAla 250
      :::::   ::   |||   :::::||||
1085  GCAGCTGGCAAGCAGCAGAGGACTCGGCCCTGTGGCTATGGCTGCACGCC 1134
      :::::   ::   |||   :::::||||
251  ThrGluSerArgThrCysAspArgProAsnCysProGlyIleGlu..... 265
      |||||:::   ::   |||   :::::||||
1135  ACCGAGACCGGTACCTGTGACACTGCGCTCTGCTGGCAGCTGAGGACAA 1184
      :::::   ::   |||   :::::||||
266  AspThrPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerG 282
      |||||:::   ::   |||   :::::||||
1185  GCACACCTTGGCCCTCCCGCAGTGAGAGTGGAGCTCTCTGGCCCGC.... 1230
      :::::   ::   |||   :::::||||
282  luGluPheAsnAlaThrLysLeuPheGluValAspThrAspSerCysGlu 298
      |||||:::   ::   |||   :::::||||
1231  .....AATGCTACGACATGATGATCAAGATGTGGACAGCTGTGAG 1272
      :::::   ::   |||   :::::||||
299  ArgTrpMetSerCysLysSerGluPheLeuLysLysTyrMetHisLysVa 315
      :::::   ::   |||   :::::||||
1273  AAGTGGCTGAACTGCAAGAGCGACTTCTCAATCAAGTATCTCAGCCAGAT 1322
      :::::   ::   |||   :::::||||
315  lMetAsnAspLeuProSerCysProCysSerTyrProThrGluValAlaT 332
      :::::   ::   |||   :::::||||
1323  GCTGGGGACCTGCGCCAGCTGCGCGTGTGCTTACCCTGAGGGCCATGG 1372
      :::::   ::   |||   :::::||||
332  yrSerThrAlaAspIlePheAspArgIleLysArgLysAspPheArgTrp 348
      |||:::   ::   |||   :::::||||
1373  ACAGCCCTGTGAGCTACAGGACGAGCACCAGGGCGCGCAGCTTCCGGTGG 1422
      :::::   ::   |||   :::::||||
349  LysAspAlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAl 365
      :::::   ::   |||   :::::||||
1423  AGGGATGCGACGTGGCCCTCGCGAGCGCCCTGGACATCTACACGCCACGGC 1472
      :::::   ::   |||   :::::||||
365  aArgTyrCysIleArgSerMetLeuSerLeuGluSerThrThrLeuAlaA 382
      |||||:::   ::   |||   :::::||||
1473  GCGCTTTCCTGCTGCTTCCATGCTGCTCTGGGGAGAGCAGCAGCTGGCCG 1522
      :::::   ::   |||   :::::||||
382  laGlnHisCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLys 398
      :::::   ::   |||   :::::||||
1523  CCCAGCACTGCTGCTATGACGAGGACAGCGGCTGCTGACCCGTGGCAAG 1572
      :::::   ::   |||   :::::||||
399  GlyAlaGlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuHi 415
      |||||:::   ::   |||   :::::||||
1573  GCGCGCGCATGCCCAACCTCATCAGCACCGCACTTCTCACCTAAGCTGCA 1622
      :::::   ::   |||   :::::||||
415  sfyrLysValAspValLeuProTrpIleCysLysGlyAspTrpSerA 432
      ||:::   ::   |||   :::::||||
1623  CTTCAAGTTGACAGCAGCGCCCTGGATCTGTGCAAGGGGAGCTGGAGCC 1672
      :::::   ::   |||   :::::||||
432  rgTyrAsnGluAlaArgProProAsnAsnGlyGlnLysCysThrGluSer 448
      ||   ::   ::   |||||:::   ::   |||   :::::||||
```

```

1673  GCCTCCAGCTGTCTCCTCCCAACAACAGCGCCGAGCTGCACCCACAAC 1722
449  ProSerAspGluAspTyrIleLysGlnPheGluAlaArgGluTyr 464
      |||:::   ::   |||   :::::||||
1723  CCCCTGGAGGAGGACTACCTAGCACAGTTGCAGGAGGCCAAGGAGTAC 1770
seq_name: gb_pat:AX351004
seq_documentation_block:
LOCUS AX351004 495 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 3 from Patent WO0190179.
ACCESSION AX351004
VERSION AX351004.1 GI:18616373
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Turner,C.A., Hilbun,E., Donoho,G., Friedrich,G.C., Abuin,A.,
Zambrowicz,B. and Sands,A.T.
TITLE Novel human thrombospondin-like proteins and polynucleotides
encoding the same
JOURNAL Patent: WO 0190179-A 3 29-NOV-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source 1..495
/organism="Homo sapiens"
BASE COUNT 126 a 156 c 139 g 74 t
ORIGIN
alignment_scores:
Quality: 900.00 Length: 164
Ratio: 5.488 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-863-824-2 x AX351004 ..
Align seg 1/1 to: AX351004 from: 1 to: 495
301 MetSerCysLysSerGluPheLeuLysLysTyrMetHisLysValMetAs 317
|||||:::   ::   |||   :::::||||
1 ATGAGCTGCAAAAGCGAGTCTTTAAAGAGATGATGACAAAGGTGATGAA 50
317 nAspLeuProSerCysProCysSerTyrProThrGluValAlaTyrSer 334
|||||:::   ::   |||   :::::||||
51 TGACCTGCCAGCTGCCCTGCTCTCTACCCCACTCAGGTGGCCTACAGCA 100
334 hrAlaAspIlePheAspArgIleLysArgLysAspPheArgTrpLysAsp 350
|||||:::   ::   |||   :::::||||
101 CGCGCGACATCTTCGACCGCATCAAGCGCAAGGAGCTTCGCTGGGAAGAC 150
351 AlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAlaArgTy 367
|||||:::   ::   |||   :::::||||
151 GCACGGGCGCCCAAGGAGAGCTGGAGATCTACAAGCCCACTGCCCGGTA 200
367 rCysIleArgSerMetLeuSerLeuGluSerThrThrLeuAlaAlaGlnH 384
|||||:::   ::   |||   :::::||||
201 CTGCATCCGCTCCATGCTGCTCCCTGGAGAGCACCACGCTGGCGGCACAGC 250
384 lsCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLysGlyAla 400
|||||:::   ::   |||   :::::||||
251 ACTGTGCTACGGCGCACAAACATGACAGCTCATCACCAGGGGCAAGGGGCG 300
401 GlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuHisTyrLy 417
|||||:::   ::   |||   :::::||||
301 GGCAGCGCCCAACCTCATCAGCACCGAGTCTCCGCGGAGCTCCACTACAA 350
417 sValAspValLeuProTrpIleCysLysGlyAspTrpSerArgTyrA 434
|||||:::   ::   |||   :::::||||
```

```
351 GGTGAGGTCCTGCTGCTGATTCCTGCAAGGGTGACTGGAGCAGGTATA 400
434 snGluAlaArgProProAsnAsnGlyGlnLysCysThrCluSerProSer 450
401 ACAGAGCGCGGCTCCACACGACGACAGAGTGACACAGAGCCCTCG 450
451 AspGluAspTyrIleLysGlnPheGlnGluAlaArgGluTyr 464
451 GACGAGGACTATCATCAGCAGTTCACAGAGGCCAGGGAATAT 492
```

```
seq_name: gb_pr:BC017997
seq_documentation_block:
LOCUS BC017997 1617 bp mRNA linear PRI 06-DEC-2001
DEFINITION Homo sapiens, clone IMAGE:4252124, mRNA, partial cds.
ACCESSION BC017997
VERSION BC017997.1 GI:17389973
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1617)
Direct Submission
Strausberg, R.
Submitted (03-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 35 Row: f Column: 10  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

Location/Qualifiers  
1..1617  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4252124"  
/tissue\_type="Skin, melanotic melanoma, high MDR."  
/clone\_lib="NIH\_MGC\_62"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"  
<1..457  
/codon\_start=2  
/product="Unknown (protein for IMAGE:4252124)"  
/protein\_id="AAH17997.1"  
/db\_xref="GI:17389974"  
/translation="KVMNDLPSCSPYFEVAYSTADIFDKRKDFRKWDAGPKK  
LEIYPTCYIRMSLESTLAAHCYGDNMQLITRGKAGTPNLITFEFAELH  
YKVDVPLTWICKDMSRYNEARPNNGKCTESPSEDYIKQFOAREY"

```
FEATURES
source
CDS
BASE COUNT 506 a 348 c 391 g 372 t
ORIGIN
```

```
alignment_scores:
Quality: 828.00 Length: 151
Ratio: 5.483 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-863-824-2 x BC017997
Align seg 1/1 to: BC017997 from: 1 to: 1617
314 LysValMetAsnAspLeuProSerCysProCysSerTyrProThrGluVa 330
|||||
2 AAGGTGATGAATGACCTGCCAGCTGCTCCCTACCCCACTGAGGT 51
330 lAlaTyrSerThrAlaAspIlePheAspArgIleLysArgLysAspPheA 347
|||||
52 GGCTACAGCAGCGCCGACATCTTCGACCGCATCAAGCAAGCAAGGACTTC 101
347 rGTpLysAspAlaSerGlyProLysGluLysLeuGluIleTyrLysPro 363
|||||
102 GCTGAAGGAGCGCCAGCGGCCCAAGGAGAAGCTGGAGATCTACAAGCCC 151
364 ThrAlaArgTyrCysIleArgSerMetLeuSerLeuGluSerThrThrLe 380
|||||
152 ACTGCCGGTACTGCATCCGCTCCATGCTGTCTCTGGAGAGCACCAGCT 201
380 uAlaAlaGlnHisCysCysTyrGlyAspAsnMetGlnLeuIleThrArgG 397
|||||
202 GCGGCACAGCAGCTGCTGCTACGGCGACACATGCAGCTCATCACCAGGG 251
397 lYlYsGlyAlaGlyThrProAsnLeuIleSerThrGluPheSerAlaGlu 413
|||||
252 GCAAGGGGGCGGCGACGCCCAACCTCATCAGCAGCGAGTTCTCCGCGGAG 301
414 LeuHisTyrLysValAspValLeuProThrIleCysLysGlyAspTr 430
|||||
302 CTCCACTACAAGGTGAGCTGCTGCCCTGGATTATCTCAAGGGTGACTG 351
430 pSerArgTyrAsnGluAlaArgProProAsnAsnGlyGlnLysCysThrG 447
|||||
352 GAGCAGGTATACAGAGGCGCGGCTCCCAACACGACAGCAAGTGACAG 401
447 LuSerProSerAspGluAspTyrIleLysGlnPheGlnGluAlaArgGlu 463
|||||
402 AGAGCGCCCTCGGACGAGGACTACATCAAGCAGTTCCTCAAGAGGCGGAA 451
464 Tyr 464
452 TAT 454
seq_name: gb_pr:AF111168
seq_documentation_block:
LOCUS AF111168 231464 bp DNA linear PRI 14-JUL-1999
DEFINITION Homo sapiens serine palmitoyl transferase, subunit II gene,
complete cds; and unknown genes.
ACCESSION AF111168
VERSION AF111168.2 GI:5468517
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 231464)
Multimegabase Sequencing Group.
Complete sequence of the gene for serine palmitoyltransferase,
subunit II found on human chromosome 14
Unpublished
2 (bases 1 to 231464)
UW Multimegabase Sequencing Group.
Direct Submission
TITLE
JOURNAL
AUTHORS
JOURNAL
COMMENT
Submitted (27-DEC-1998) Department of Molecular Biotechnology, Box
35730 University of Washington, Seattle, Washington 98195, USA
On Jul 14, 1999 this sequence version replaced gi:4186181.
Sequencing methodology: high redundancy shotgun with plasmids.
Interspersed Repeats were identified with RepeatMasker (available
from http://ftp.genome.washington.edu/RN/RepeatMasker.html).
```

FEATURES	Location/Qualifiers	
source	1. .231464	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="14"	
	/map="14q24.3-31"	
	/clone="BAC 50114"	
	/clone_lib="RPCI-11"	
	9. .186	
repeat_region	/rpt_family="L1MD1"	complement(9631..9708)
repeat_region	200..494	/rpt_family="LINE2"
repeat_region	/rpt_family="AluSx"	complement(10461..10762)
repeat_region	500..1283	/rpt_family="AluSx"
repeat_region	/rpt_family="AluSg"	complement(10804..11077)
repeat_region	1284..1581	/rpt_family="AluJb"
repeat_region	/rpt_family="AluSg"	complement(11156..11306)
repeat_region	1582..1774	/rpt_family="MIR"
repeat_region	/rpt_family="AluJb"	complement(11493..11615)
repeat_region	/rpt_family="AluJb"	/rpt_family="FLAM_C"
repeat_region	1775..2058	11800..12161
repeat_region	2059..2934	/rpt_family="THE1C"
repeat_region	/rpt_family="AluJb"	complement(12406..12695)
repeat_region	complement(2937..3255)	/rpt_family="AluSx"
repeat_region	/rpt_family="MER4A"	12746..12920
repeat_region	323..3687	/rpt_family="BC200"
repeat_region	/rpt_family="L1"	12921..13217
repeat_region	3688..3977	/rpt_family="AluSx"
repeat_region	/rpt_family="AluSg"	13236..13751
repeat_region	4009..4314	/rpt_family="L1ME1"
repeat_region	/rpt_family="AluJo"	13783..14084
repeat_region	4321..4934	/rpt_family="AluSx"
repeat_region	/rpt_family="L1"	14085..14403
repeat_region	4940..5236	/rpt_family="L1ME2"
repeat_region	/rpt_family="AluSp"	14405..14717
repeat_region	5238..5349	/rpt_family="AluSx"
repeat_region	/rpt_family="L1MC1"	14795..14857
repeat_region	5350..5650	/rpt_family="L1MC2"
repeat_region	/rpt_family="AluSg"	complement(14884..14905)
repeat_region	5668..5924	/rpt_family="AT-rich"
repeat_region	/rpt_family="AluSg"	complement(15106..15393)
repeat_region	5669..5940	/rpt_family="AluSx"
repeat_region	/rpt_family="CAAAA)n"	complement(15606..15818)
repeat_region	6086..6271	complement(15835..16131)
repeat_region	/rpt_family="AluJb"	/rpt_family="AluSc"
repeat_region	complement(6500..6588)	complement(16775..17073)
repeat_region	/rpt_family="5S"	/rpt_family="AluY"
unsure	complement(6715..6815)	17491..17777
repeat_region	/note="low quality data"	/rpt_family="AluJo"
repeat_region	complement(6914..7458)	17778..17804
repeat_region	/rpt_family="LINE2"	/rpt_family="(CA)n"
repeat_region	7523..7707	complement(17815..17882)
repeat_region	/rpt_family="L1MB6"	/rpt_family="L1MB2"
repeat_region	7708..8007	17883..18175
repeat_region	/rpt_family="AluSx"	/rpt_family="AluSx"
repeat_region	8233..8305	18177..18251
repeat_region	/rpt_family="(TA)n"	/rpt_family="(TAGA)n"
repeat_region	9186..9230	complement(18244..18306)
CDS	complement(join(9248..9470,10060..10240,16529..16722,25330..25484))	/rpt_family="L1MB1"
	/note="Intron-exon boundaries defined in relation to an EST contigs that includes AA397580, A1138314, AA399545, AA868184, all from testis. The closest match in BLASTX is to E. coli pyrroline-5-carboxylate reductase."	complement(18310..18609)
	/codon_start=1	/rpt_family="AluSx"
	/product="unknown"	complement(18624..18771)
	/db_xref="GI:4186186"	/rpt_family="L1MB3"
	/translation="MDMLQDLESLOFEXGVPEEDRIWLVLQGRSLGRLTEACAHATFFCKLLYNLRASLNKNSRHLSIGLSNATPEEFKVCIIIGGHLGKLAGTLLQLGPILASRLISRRPTLGLQKIGKCFYHNADLVSWADVIFLCLLPOLPNICVEITSL EKASIVYFVAAPILPRLLNLTNLRPQYDSDSVWGNKGVTAALQDPTIL QATCPSAGHSGTYFWHSFENYFPVGPLLSY"	complement(18765..18847)
		/rpt_family="L1"
		18858..19501
		/rpt_family="SVA"
		complement(19503..20417)
		/rpt_family="L1"
		complement(20104..20625)
		/rpt_family="L143_Send"
		complement(20724..21029)
		/rpt_family="AluJb"
		complement(21030..21122)
		/rpt_family="(TAAA)n"
		complement(21125..21423)
		/rpt_family="AluJo"
		complement(21486..21662)
		/rpt_family="L191_Send"
		complement(21664..21966)
		/rpt_family="AluSg"
		complement(22042..22342)

```
repeat_region      /rpt_family="AluY"
complement(22350..22937)
/rpt_family="L143_5end"
repeat_region      complement(22975..23090)
/rpt_family="FLAM_C"
repeat_region      complement(23164..23287)
/rpt_family="L143_5end"
repeat_region      complement(26429..26704)
/rpt_family="L1MB5"
repeat_region      26766..27080
/rpt_family="AluX"
repeat_region      complement(28232..28432)
/rpt_family="MER20"
repeat_region      28528..28689
/rpt_family="AluJb"
repeat_region      28692..28984
/rpt_family="AluX"
repeat_region      complement(29016..29108)
/rpt_family="MIR"
CDS
complement(join(30211..30231,30965..31069,31601..31690,
32276..32362,36437..36526,36857..36898,37854..37988,
38439..38514,40403..40476,43661..43688,45155..45257,
45714..45747,46844..46936,51089..51145,51888..51952,
53782..53928,55894..55996,56605..56697))
note="Intron-exon boundaries defined by an EST contig
that includes W3352, W80307, AA780270, AA278570, R53310,
and AA437679. There is a short form of this coding
sequence that terminates at position 188433, based on EST
AA780270. The closest similarity by BLASTX is to a C.
elegans hypothetical protein."
```

## alignment\_scores:

```
Quality: 634.00      Length: 172
Ratio: 4.227         Gaps: 0
Percent Similarity: 87.209 Percent Identity: 60.465
```

## alignment\_block:

US-09-863-824-2 x AF111168/rev ..

Align seg 1/1 to reverse of: AF111168 from: 1 to: 231464

```
293 AspThrAspSerCysGluArgTrpMetSerCysLysSerGluPheLeuIys 309
78708 GATGTGACAGCTGTGAGAAGTGGTGAACACTGCAAGAGCGACTTCTTAAT 78659
309 sLysTyMetHisLysValMetAsnAspLeuProSerCysProCysSerT 326
78658 CAAGTATCTGAGCCAGATGCTGGGGACCTGCCAGCTGCCCGTGTGCCT 78609
326 yrProThrGluValAlaLysSerThrAlaAspIlePheAspArgIleLys 342
78608 ACCCACTGGAGGCCATGGACAGCCCTGTGAGCCTACAGGACGAGCACAG 78559
343 ArgLysAspPheArgTrpLysAspAlaSerGlyProLysGluLysLeuCl 359
78558 GGCCGCGAGCTTCGGTGGAGGATGCACTGGCCCTCGCGAGCGCGCTGGA 78509
359 uileTyLysProThrAlaArgTyrCysIleArgSerMetLeuSerLeuG 376
78508 CATCTACCAACCCACCGCGGCTTCTGCTGCTCCATGCTGTCTGGGG 78459
376 luSerThrThrLeuAlaAlaGlnHisCysCysTyrGlyAspAsnMetGln 392
78458 AGAGCAGCACACTGGCCGCCACGACTGCTGTATGACGAGGACAGCCGG 78409
393 LeuIleThrArgGlyLysGlyAlaGlyThrProAsnLeuIleSerThrGl 409
78408 CTGCTGACCGTGGCAAGGCGCGCATGCTCCCAACCTCATCAGCACCGA 78359
409 uPheSerAlaGluLeuHisTyrLysValAspValLeuProTrpIleIleC 426
78358 CTTCCTACCTAAGCTGCACTTCAGTTCGACACGACGCGCCCTGGATCCTGT 78309
```

```
426 ysLysGlyAspTrpSerArgTyrAsnGluAlaArgProAsnAsnGly 442
78308 GCAAGGGGACGTGGAGCGCTCCACGCTGTGCTCCCTCCCAACACGCG 78259
443 GlnLysCysThrGluSerProSerAspGluAspTyrIleLysGlnPheGl 459
78258 CGAGCCTGCACCGACAAACCCCTGGAGGAGGAGTACCTAGCACAGTTGCA 78209
459 nGluAlaArgGluTyr 464
78208 GGAGGCCAAGGAGTAC 78193
seq_name: gb_pr:HSU107712
```

## seq\_documentation\_block:

```
LOCUS      HUMAN DNA sequence from clone RP5-107712 on chromosome 20. Contains
DEFINITION part of a novel gene, ESTs, STSs, GSSs and a CpG island, complete
sequence.
ACCESSION  AL050320
VERSION    AL050320.19 GI:7263998
KEYWORDS   HTG: CpG island.
SOURCE     human.
ORGANISM   Homo sapiens
```

```
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS    Smith,M.
TITLE      Direct Submission
JOURNAL    Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
            requests: clones@sanger.ac.uk
```

## COMMENT

On Mar 19, 2000 this sequence version replaced gi:6523711.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
EM: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 20, constructed by the Sanger Centre Chromosome 20  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr20

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e. phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. RP5-107712 is from  
the library RPCI-5 constructed by the group of Pieter de Jong. For  
further details see  
http://www.chori.org/bacpac/home.htm

## FEATURES

source Location/Qualifiers
 1..107057
 /organism="Homo sapiens"
 /db\_xref="taxon:9606"
 /chromosome="20"
 /clone="RP5-107712"
 /clone\_lib="RPCI-5"
 repeat\_region 286..490

## FEATURES

source Location/Qualifiers
 1..107057
 /organism="Homo sapiens"
 /db\_xref="taxon:9606"
 /chromosome="20"
 /clone="RP5-107712"
 /clone\_lib="RPCI-5"
 repeat\_region 286..490





Fri Apr 26 10:30:00 2002

```
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-863-824-2 x HSJ107712 ..
Align seg 1/1 to: HSJ107712 from: 1 to: 107057
47 AsnAsnLeuAsnValGlySerAspThrThrSerGluThrSerPheSerLe 63
|||||
106877 AATAACCTCAACGTGGGAAGTGACACCATCAGAAACACGCTTTCTCT 106926
63 uSerLysGluAlaProArgGluHisLeuAspHisGlnAlaHisGlnP 80
|||||
106927 CTCRAAGAAGCACCAGGAGCATCTGGACCACCGAGCTGCACACCAAC 106976
80 roPheProArgProArgPheArgGlnGluThrGlyHisProSerLeuGln 96
|||||
106977 CCTTCCCGACAGCGGATTCGACAGAGACGGGACCCCTTCATTCGAA 107026
97 ArgAspPheProArgSerPheLeuLeuAsp 106
|||||
107027 AGAGATTTCCCGAGATCCTTCTCTTGAT 107056
seq_name: gb_sts:HS107712S
seq_documentation_block:
LOCUS HS107712S 523 bp DNA linear STS 24-AUG-1999
DEFINITION H.sapiens STS from genomic clone 107712, sequence tagged site.
ACCESSION AL110084
VERSION AL110084.1 GI:5777434
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 523)
AUTHORS Deloukas,P., O'Neill,L., Holden,J., Mistry,D., Huckle,E., Taylor,R.
and Hunt,S.
Direct Submission
Submitted (24-AUG-1999) E-mail contact: humquery@sanger.ac.uk
Marker: STJ107712SP6 (Primer A : GGAATCTCTTTCATGAGG; Primer B :
ACCTCAACGTGGGAAGTAC; amplicmer size : 156 bp) is from sequence
generated from the Sp6 end of PAC 107712. 107712 is part of the
bacterial clone contigs constructed by the Chromosome 20 Mapping
Group.(http://www.sanger.ac.uk/HGP/Chr20/) 107712 is from the
library constructed at the Roswell Park Cancer Institute by the
group of Pieter de Jong.
For further details see http://bacpac.med.buffalo.edu/.
FEATURES
Location/Qualifiers
1..523
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="107712"
BASE COUNT 150 a 102 c 116 g 155 t
ORIGIN
alignment_scores:
Quality: 292.00 Length: 54
Ratio: 5.407 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-863-824-2 x HS107712S/rev ..
Align seg 1/1 to reverse of: HS107712S from: 1 to: 523
47 AsnAsnLeuAsnValGlySerAspThrThrSerGluThrSerPheSerLe 63
|||||
162 AATAACCTCAACGTGGGAAGTGACACCATCAGAAACACGCTTTCTCT 113
63 uSerLysGluAlaProArgGluHisLeuAspHisGlnAlaHisGlnP 80
|||||
```

```
|||||
112 CTCCTCAAGAGCACCAGGAGCATCTGGACCACCGAGCTGCACCAAC 63
80 roPheProArgProArgPheArgGlnGluThrGlyHisProSerLeuGln 96
|||||
62 CCTTCCCGACAGCGGATTCGACAGAGACGGGACCCCTTCATTCGAA 13
97 ArgAspPhePro 100
|||||
12 AGAGATTTCCCC 1
seq_name: gb_ro:MUSTSP1A
seq_documentation_block:
LOCUS MUSTSP1A 4339 bp mRNA linear ROD 09-NOV-1994
DEFINITION Mouse thrombospondin 1 mRNA, complete cds.
ACCESSION M87276
VERSION M87276.1 GI:202196
KEYWORDS thrombospondin 1.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4339)
AUTHORS Laherty,C.D., O'Rourke,K., Wolf,F.W., Katz,R., Seldin,M.F. and
Dixit,V.M.
TITLE Characterization of mouse thrombospondin 2 sequence and expression
during cell growth and development
J. Biol. Chem. 267 (5), 3274-3281 (1992)
MEDLINE 92147683
FEATURES
Location/Qualifiers
1..4339
/organism="Mus musculus"
/strain="CD-1"
/db_xref="taxon:10090"
/tissue_type="kidney"
/dev_stage="embryo"
212..3724
/codon_start=1
/product="thrombospondin 1"
/protein_id="AA53063.1"
/db_xref="GI:567240"
/translat="MELLRLGLVFLHLMCGSNRIPESGGDNGVDFIFELIGARRGP
GRPLVKGDLSSPAFENANLIPVDDKFODLLDAVMADKGFIFLASLRQMKKTRG
TLIAVERKNTQIFSVNSGKAGTLDLSLPGKQVVSVVEALLATQOMKSLTFV
QEDRAVLLIDCKMESAEADVFIQSFTRDLASVARLVAKGVDVNDQGVQLQNVREV
FGTTPEDILRNKCSSTNVLITLDNNVNGSPAIRTNVIGHKTKDLQAIGLSCDE
LSSMVLKGLRTIVTTLQDSIRKVTENRELVSELKRPPLCFHNGVDTKNNEMTVD
SCTECHQNSVTICKKVCPIMPCSNATVPDGECCPCRPWSDSADGGVSPSWSCSVT
ATCGNGIQGRSCDSLNNRCSSVQTRTCHTQCDKRFKQDGGWSHSPWSDSVT
CDGVVTRILCNPSPOMGKPCGEARETRACKKDCACPIDGCLSNPCFAGAKTSYV
GGVQRRSLCNPTQFGGKDCQDVTEQVNCNKDCCPIDGCLSNPCFAGAKTSYV
DGSWKCGCPGPGYSGNGIQCQDVDECKEVPDCAFNHNGHRCNKTDGYNCLPCPPRF
TGSOPEGRGVEHAMANKQVCKPRNCTDGHDCNKNKAKNYLGHVSQDPMYRCCKPGY
AGNGIICGEDTDLGWPENLVANATYHCKKDCNPNLPSGQEDYDKDGDGACDD
DDNDKIIPDDRNCQPHYNPAQYDTRDDVDGDCNCPYHNHPDQADTKNGEGDACA
VDIDGGLINERDNCQYVNVQDRTDMDGVGDCQDCNCPLEHNPQDLGSLDGTG
DNNQDIDEGHQNHLNCPVYVNAOAHDKGKDACDHDNDGIPDDRNCRLVP
NPDKSDSDGGRGACKDQDVNDVDPIDDIICPENFDISETDPRRFQMLPLDPKGTQ
NDPNWVRVGRGKELVQTCPLGLAVDFIDFNADVDSGTFFINTERDDYAGLGVGQ
SSSRFYVVMKQVTSYDTPNTRAQYSGLSVKVYVNSTTGPGEHLRNALHMTGNTPG
QVRTLWHDPRHIGWKDFTYRWRLSHRPKTGYIRVVMYEGKKIMADSGPIYDKTYAGG
RLGLVFVSQEMVFFSDMKYBCRDS"
BASE COUNT 1109 a 1201 c 1144 g 885 t
ORIGIN
alignment_scores:
Quality: 145.50 Length: 328
Ratio: 0.951 Gaps: 17
Percent Similarity: 46.646 Percent Identity: 24.085
alignment_block:
```

```

US-09-863-824-2 x MUSTSP1A ..
Align seg 1/1 to: MUSTSP1A from: 1 to: 4339

15 LeuLeuLeuThrLeuHisIleThrValLeuArgGlySerGlyAlaAla.A 31
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
923 GTCCCTTCTTACCTTGACAACAAGCTGTGAACGGTTCAGCCCTCGCTAT 972

31 spGlyProAspAlaAlaGlyAsnAlaSerGlnAlaGlnLeuGlnAsn 47
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
973 CCGCACCAACTACATCGGCC.....ACAAACAAGGACCTCC 1010

48 AsnLeuAsnValGlySerAspThrThrSerGluThrSerPheSerLeu.S 64
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1011 AAGCTATCTGTGGCTCTCCTGTGATGAATATCCAGCATGCTCTGGAA 1060

64 erLysGluAlaProArgGluHisLeuAspHisGlnAlaAlaHisGlnPro 80
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1061 CT...GAAGGGCTCGGCACCATCTGTACCACCTCTGCAGGACAGCATCCG 1107

81 PheProArgProArgPheArgGln.....GluThrGlyHi 92
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1108 AAAAGTGACGAAGAACAAGAGAGCTGCTGAGTGAGCTGAAGCGGCCTC 1157

92 sProSerLeuGlnArg.....:::.....AspP 99
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1158 CCCTCTGCTTTCACAATGGAGTCCAGTACAAACAACAGGAGGTGACT 1207

99 heProArgSerPheLeuLeuAspLeuPro.AsnPheProAspLeuSerLy 115
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1208 GTAGACAGTTCACACAGAGTGTCACTGCCAGAACTCGGTTACCATCTGCAA 1257

115 sAlaAspIleAsnGlyGln**ProAsnIleGlnValThrIleGluValV 132
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1258 AAAGTGTCTGTCCCATCATCGCTGCTCCCAAGCCACAGTT..... 1300

132 alAspGlyProAspSerGluAlaAspLysAspGlnHisProGluAsnLys 148
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1301 .....CCTGATGGTGAATCC.....TGCCACGGGTGCTGG 1330

149 ProSerTrpSerValProSerProAspTrpArgAlaTrpTrpGlnArgSe 165
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1331 CCCAGCGACTGTCT...GACGATGGTGGTCTCCCTGCTGTGAGTGGAC 1377

165 rLeuSerLeuAlaArgAlaAsnSerGlyAspGlnAsp.....Tyr***T 180
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1378 CTCCTGTCTGCTGCACATGTGCAATGGAAATTCAGCAACGTGCTGCTTCT 1427

180 yrAspSerThrSerAsp.....:::..... 185
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1428 GTGACAGCCTCAACAACAGATCGGAGGGCTCTTCGGTACAGCAGGACC 1477

186 .....:::.....AspSerAsnPheLeuAsnProProArgGlyTr 196
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1478 TGCCACATTCAGGAGTGTCAAAAAGATT...AAACAGGATGTGTGGCTG 1524

196 pAspHisThrAlaProGlyHisArgThrPheGluThr..... 208
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1525 GAGTCATCTGGTCTCCATGGTCTCTCTGTGTGACCTGTGTGTGACGGTG 1574

209 .....:::.....LysAspGlnProGluThrAspSer 216
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1575 TGATCACAAGGATCGTCTCTGCAACTCCCCCAGCCCCCATGATGAACGG 1624

217 ThrAspGlyGluGlyAsp.....:::..... 222
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1625 AAGCCCTGTGAAGGTGAAGCCGGGAGACCAAGCCTGCAAGAAAGACGC 1674

223 .....:::.....TrpSerLeuTrpSerValCysS 230
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1675 CTGCCCAATTATGAGGCTGGGGTCCCTGTGTGTCACCATGGGACATCTGCT 1724

230 erValThrCysGlyAsnGlyAsnGlnLysArgThrArgSerCys..... 244

```

```

1725 CTGTCACTGTGGAGGAGTGCAGACAGCCGACCTGTGTAACAAC 1774
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
245 .....GlyTyrAlaCystThr.....AlaThrGluSe 253
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1775 CCCACACCCCAAGTTGGAGGCAAGACTGTGTGGCGATGTGACAGAAA 1824

253 rArgThrCysAspArgProAsnCysPro 262
   ::::::::::::::::::::: ::::::::::::::::::::: ::::
1825 TCAAGTTTGCACAAAGCAGGACTGCCCA 1852

seq_name: gb_htg:AL626774

seq_documentation_block:
LOCUS AL626774 199982 bp DNA linear HTG 21-OCT-2001
DEFINITION Mus musculus chromosome 4 clone RP23-198015, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL626774
VERSION AL626774.8 GI:16304784
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 199982)
AUTHORS Burton,J.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 21, 2001 this sequence version replaced gi:16215137.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BM198015
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 194180 bases at least Q40
Consensus quality: 195534 bases at least Q30
Consensus quality: 196426 bases at least Q20
Insert size: 197682; sum-of-contigs
Quality coverage: 8.11x in Q20 bases; agarose-fp
coverage: 8.43x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source
1. .199982
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone_lib="RPC1-23"
1. .8011
/notes="assembly_fragment:00065
fragment_chain:1
clone_end:SP6
vector_side:left"
8112. .15629
/notes="assembly_fragment:01193
fragment_chain:1"
15730. .19452
/notes="assembly_fragment:04957
fragment_chain:1"
19553. .21702

```



SOURCE Homo sapiens brain cDNA to mRNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 5399)  
 AUTHORS Nakamura, Y.  
 JOURNAL Direct Submission  
 Submitted (28-JUN-1997) Yusuke Nakamura, The Inst. of Medical  
 Science, The University of Tokyo, Lab. of Molecular Medicine, Human  
 Genome Center, 4-6-1, Shirokanedai Minato-ku, Tokyo, Minato-ku,  
 Tokyo 108, Japan (E-mail:yusuke@ims.u-tokyo.ac.jp,  
 Tel:81-3-5449-5372, Fax:81-3-5449-5433)  
 REFERENCE 2 (slices)  
 Shiratsuchi, T., Nishimori, H., Ichise, H., Nakamura, Y. and Tokino, T.  
 Cloning and characterization of BAI2 and BAI3, novel genes  
 homologous to brain-specific angiotensin inhibitor 1 (BAI1)  
 Cytogenet. Cell Genet. 79 (1-2), 103-108 (1997)

Location/Qualifiers

1..5399  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="lp35"  
 /tissue\_type="brain"  
 379..5097  
 /gene="BAI 2"  
 379..5097  
 /gene="BAI 2"  
 /codon\_start=1  
 /protein\_id="BAA25362.1"  
 /db\_xref="gi:3021699"  
 /translation="MTAPCLLLSVLSRLATAPDAPSPACSAASGLVLYGAFSLQD  
 LFPTIASCSTLENPDPTKSYLRFNRBOVCAHFAPELLPLDHYLVNFTCLRPSP  
 EEVAQAASEVGRPEEEAEAAAGLELCSGSPTEFLHDFKNFVQLCLASBPSEAPRL  
 LAPALAFPFVEVLINNNSSQFTGVCIRWSECEGAGRAACGAFPGSCGPEAG  
 AGSTTTSPFPAATLSNALVPGGPAPPAEADLHSGSNDLFTTEMRYGPEEPK  
 VTKQWPSADPGLYMAQTDPAAEWSPVSCSLTCCGQVTRSCVSPKPTLCS  
 GLRETRPCNNASATCPVHGWEEMGWSLCSRCGSRGSRMRCTVPPHGGKCEGP  
 ELQTKLCSMAACPVGOWLENGPWGCPSCANGTQOORRKC SVAGPAWACTGALT  
 TRECNECPATDSKWGPWNAWSLCSKTDGTGWQRRFRKCATGTGQYPCGEGTGEVK  
 PCSERCAFHEMCRDEYVLMWTKAAAGELIYNKPPNAGSSASRLASLAQSVAY  
 WGLPFCFARISHEYRYLYSLREHLAKORLAGEMSQVRSQELRLARTYVSGDL  
 LFSVDILRNVDITFRATYVPSADVDQRFQVFMVDAENKEKWDQAQQVSPGSHL  
 RVVEDFLHVGDAKAFQSSLIYTDNLVISTOREPVSAVSDITFPMRGRGMKDW  
 RHSEDRFLPAXEVLSLSPGKPAATSGAAGPCRGCTVPPGCHSHQRLLPADPE  
 SSYFVIGAVLRTGLILPPRPPLAVTSRVTVTPPTQPPAEPLITVELSYIING  
 TTDPHCASDYSRADASSGDWDTENCQTLTOAAHTRCQCQHLSTFAVLAQPPKDLTL  
 ELAGSPVPLVIGCAVSCMALTLTLLAYAFWRFITKRSRIILNFCILSLASNLIL  
 VQSRVLSKGVCTMTAFLHFFLSFCWLTAEWQSYLAVIGRMRTLVRKRFLCLG  
 WGLPALVVASVGTGRTKGYGTSVYCWLSLEGLLYAFVGAATVILVNLGILIVFN  
 KLMDGIDSRSKQKORAGSERCPWASLLIPCSAGCAVPSPLSSASARNASLWSSC  
 KLMARDGLTWSAVLMDRRSVLFQALFAVNSAGFEVITAVHFLRREVODVVYKQ  
 MGVCLADESDPSCKNGQLQILSDFEKDVLACQTVLFKEVNTCPNSTITGLSLR  
 SLDEDEEPKCSLGVPEGLSPFNGLNLPMAASPGLEPPPOEANPVMCMGSGGL  
 ROLDLTLPEKPEGSEGDMVLPRLTSLQPGGGGGEDAPRARPECTPRAAKTVA  
 HTEGYPSLVDHSLGLGIPAYGSLQNPYGMTFPPPTPSAROVPEGERSRTPRT  
 VPGSTMKGSLERKRLRYSDLDPEVMTFRKHSLEYHLNOKFHFDFRYSQSTAKRE  
 KRWSVSSGGAERSVCTDKPSPGERPSLSQHRHRSQSWSTFKSMTLGSLPPKRLTL  
 HRAAAWEPTPPDGDQFEV"

polya\_site  
 BASE COUNT 930 a 1790 c 1652 g 1027 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 143.00 Length: 457  
 Ratio: 0.745 Gaps: 19  
 Percent Similarity: 42.013 Percent Identity: 19.912

alignment\_block:  
 US-09-863-824-2 x AB005298 ..  
 114 SerLysAlaAspIleAsnGlyGln\*\*ProAsnIleGlnValThrIleG1 130  
 1120 GCTGAGGCCGATTTGCACCTGGGAGGAGCAGCATGCTGTTTCAACACCGA 1169  
 130 uValValAspGlyProAspSerGluAlaAspLysAspGlnHisProGluA 147  
 1170 GATGAGATATGGTGAGGAGCGGAA.....GAGGAACCGAAAG 1207  
 147 snLysProSerTrp.....SerValProSerPro..... 156  
 1208 TGAACACCCAGTGGCCGAGGTCTGCAGATGAGCCTGGGCTATACATGGCG 1257  
 157 .....AspTrpArgAlaTrpTrpGlnArgSe 165  
 1258 CAGACAGCGACCGCGCGGTGAGGAGTGTGCCCGTGGAGCGGTGTGTTTC 1307  
 165 rLeuSerLeuAlaArgAlaAsnSerGlyAspGlnAspTyr\*\*TyrAsps 182  
 1308 CTTGACG.....TGTGGCAGGTCTGCAGGTGGGACCGCTCCT 1348  
 182 erThrSerAspAspSerAsnPheLeu.....Asn 191  
 1349 GTGTGTCTCTCCCTATATGGACCTGTGTGCAGCGGGCCCTCGGGGAGACC 1398  
 192 ProProArgGlyTrpAspHisThrAlaProGlyHisArgThrPheGlu... 207  
 1399 AGGCCCTGCAACAATTCAGCCACCTGCCACGTGCACGCGGTGTGGGAGA 1448  
 207 ..... 207  
 1449 GTGGGGTCTCTGGAGCCTGTCTCCCGCAGCTGCGGGGGGTCCCGGA 1498  
 208 .....ThrLysAspGlnProGluTyrAspSerThrAspGlyGlu 220  
 1499 GCGCGATGCGGACCTGTGTCGCCCGCCAGCAGCGCGGAGCGCTGCGAG 1548  
 221 Gly..... 221  
 1549 GGTCTGAGCTGCAGACATAGCTGTGCAGTATGCTGCCCTGCCCGGTGGA 1598  
 222 .....AspTrpSerLeuTrpSerValCysSerValThrCysG 234  
 1599 AGGCCAGTGGTTAGAAATGGGGTCTCTGGGGCCCATGCTCCACGCTCTGTG 1648  
 234 lyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAla..... 247  
 1649 CCAATGGGACCAACAGCGAGCGGAAGTGCAGCGTGGCGGGCCAGCC 1698  
 248 .....CysThrAla.....ThrGluSerArgThrCysAspArgPr 259  
 1699 TGGGCCACATGACGGGTGCCCTCCTCCTGACCCGGGAGTGCAGCAACCT 1748  
 259 oAsnCysProGlyIleGluAspThrPheArgThrAlaAlaThrGluValS 276  
 1749 CGAGTGGCGGCCACTGATACCAAGTGG..... 1776  
 276 erLeuLeuAlaGlySerGluGluPheAsnAlaThrLysLeuPheGlu... 291  
 1777 .....GGGCCATGGATGCTGAGCGCTGTGCTCTAAG 1809  
 292 ...ValAspThrAspSerCysGluArgTrpMetSerCysLysSerGluPh 307  
 1810 ACGTGTACACAGCGCTGCAGCGCGCTTCCGCGCATGTGCCAGGCGACGG 1859  
 307 eLeuLysLysTyrMethIleLysValMetAsnAspLeuProSerCysProC 324  
 1860 CACCGAGGGCTAC.....CCCT 1876  
 324 ysSerTyrProThrGluValAlaTyrSerThrAlaAsp..... 336  
 1144 SerLysAlaAspIleAsnGlyGln\*\*ProAsnIleGlnValThrIleG1 130  
 1120 GCTGAGGCCGATTTGCACCTGGGAGGAGCAGCATGCTGTTTCAACACCGA 1169  
 130 uValValAspGlyProAspSerGluAlaAspLysAspGlnHisProGluA 147  
 1170 GATGAGATATGGTGAGGAGCGGAA.....GAGGAACCGAAAG 1207  
 147 snLysProSerTrp.....SerValProSerPro..... 156  
 1208 TGAACACCCAGTGGCCGAGGTCTGCAGATGAGCCTGGGCTATACATGGCG 1257  
 157 .....AspTrpArgAlaTrpTrpGlnArgSe 165  
 1258 CAGACAGCGACCGCGCGGTGAGGAGTGTGCCCGTGGAGCGGTGTGTTTC 1307  
 165 rLeuSerLeuAlaArgAlaAsnSerGlyAspGlnAspTyr\*\*TyrAsps 182  
 1308 CTTGACG.....TGTGGCAGGTCTGCAGGTGGGACCGCTCCT 1348  
 182 erThrSerAspAspSerAsnPheLeu.....Asn 191  
 1349 GTGTGTCTCTCCCTATATGGACCTGTGTGCAGCGGGCCCTCGGGGAGACC 1398  
 192 ProProArgGlyTrpAspHisThrAlaProGlyHisArgThrPheGlu... 207  
 1399 AGGCCCTGCAACAATTCAGCCACCTGCCACGTGCACGCGGTGTGGGAGA 1448  
 207 ..... 207  
 1449 GTGGGGTCTCTGGAGCCTGTCTCCCGCAGCTGCGGGGGGTCCCGGA 1498  
 208 .....ThrLysAspGlnProGluTyrAspSerThrAspGlyGlu 220  
 1499 GCGCGATGCGGACCTGTGTCGCCCGCCAGCAGCGCGGAGCGCTGCGAG 1548  
 221 Gly..... 221  
 1549 GGTCTGAGCTGCAGACATAGCTGTGCAGTATGCTGCCCTGCCCGGTGGA 1598  
 222 .....AspTrpSerLeuTrpSerValCysSerValThrCysG 234  
 1599 AGGCCAGTGGTTAGAAATGGGGTCTCTGGGGCCCATGCTCCACGCTCTGTG 1648  
 234 lyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAla..... 247  
 1649 CCAATGGGACCAACAGCGAGCGGAAGTGCAGCGTGGCGGGCCAGCC 1698  
 248 .....CysThrAla.....ThrGluSerArgThrCysAspArgPr 259  
 1699 TGGGCCACATGACGGGTGCCCTCCTCCTGACCCGGGAGTGCAGCAACCT 1748  
 259 oAsnCysProGlyIleGluAspThrPheArgThrAlaAlaThrGluValS 276  
 1749 CGAGTGGCGGCCACTGATACCAAGTGG..... 1776  
 276 erLeuLeuAlaGlySerGluGluPheAsnAlaThrLysLeuPheGlu... 291  
 1777 .....GGGCCATGGATGCTGAGCGCTGTGCTCTAAG 1809  
 292 ...ValAspThrAspSerCysGluArgTrpMetSerCysLysSerGluPh 307  
 1810 ACGTGTACACAGCGCTGCAGCGCGCTTCCGCGCATGTGCCAGGCGACGG 1859  
 307 eLeuLysLysTyrMethIleLysValMetAsnAspLeuProSerCysProC 324  
 1860 CACCGAGGGCTAC.....CCCT 1876  
 324 ysSerTyrProThrGluValAlaTyrSerThrAlaAsp..... 336

1877	GCAGGGCCACCGGAGGAGGTGAAGCCCTTGTTAGTGAGAAGAGGTTC	CA	1926		
337	IlePheAspArgIleLysArgLysAspPhe.....ArgTirpLy	349			
	: : : : :     : : : : :				
1927	GCTTCATGAGTATGCAGGGATGAGTACGTGATCCTGATGACGTGGAA	1976			
349	sAspAlaSerGlyProLysGluLysLeuGluIleTyrlsProThrAla.	365			
	: : : : : : : : : : : : : : :				
1977	GAAGGCAGCTGCTGGCGAGATCATCTACACAAGATGCCCCCCGAATGCCT	2026			
366	.....ArgTyrCysIleArgSer.....	371			
	: : : : :				
2027	CAGGGTCTGCCAGCCCGCGCTGTCTCCTCAGTCCCACAGCGTGGCGTAC	2076			
371	.....	371			
2077	TGGGGGCTGCCAGCTTTGCTGCTGCATCTCCATGATACGCTACCT	2126			
372	.MetLeuSerLeuGluSerThrThrLeuAlaAlaGlnHisCysCysTyrG	388			
	: : : : : : : : : : : : : : :				
2127	GTATCTGTCACCTTAGGAGCACCTGGCCAAGGGCAGCGCATGCTGGCAG	2176			
388	lyAspAsnMetGlnLeuIleThrArgGlyLysGlyAlaGlyThrProAsn	404			
	: : : : : : : : : : : : : : :				
2177	GCAGGGCATGTCGAGGTGGTGGC.....AGCCTGCAGGAG	2214			
405	.LeuIleSer.....ThrGluPheSerAlaGluLeuHistyrlsValas	419			
	: : : : :     : : : : :     : : : : :				
2215	CTACTGCCCGGGCACCCTACTACTAGTGGGACCTGCTCTCTGTGGGA	2264			
419	pValLeuProTrpIleilecysLysGlyAspTirpSerArgTyrAsnGluA	436			
	: : : :				
2265	CAITCTG.....	A 2272			
436	laArgProProAsnAsnglyGlnLysCystThrGluSerProSerAspGlu	452			
	: : : : : : : : : : : : : : :				
2273	GGAATGTCACGTACACCTTTTAAGAGGGCCACCTACGTGCCTCGCTGAT	2322			
453	AspTyrIleLysGlnPheGln	459			
	: : : : :				
2323	GATGTGAGCGCTTCTTCCAG	2343			
<hr/>					
seq_name: gb_pat:E21623					
seq_documentation_block:					
LOCUS	E21623	5412 bp	DNA linear PAT 07-FEB-2001		
DEFINITION	Human BAI gene and utilization thereof.				
ACCESSION	E21623				
VERSION	E21623.1	GI:13023633			
KEYWORDS	JP 1999032766-A/2.				
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 5412)				
AUTHORS	Ryuji,T.Y.N.N.				
TITLE	Human BAI gene and utilization thereof				
JOURNAL	Patent: JP 1999032766-A 2 09-FEB-1999;				
	OTSUKA PHARMACEUT CO LTD				
COMMENT	OS Unidentified				
	PN	JP 1999032766-A/2			
	PD	09-FEB-1999			
	PF	16-JUN-1997			
	PR	JP 1997176485			
	PI	RYUJI TOKINO,YUSUKE NAKAMURA			
	PC	C12N15/09,A61K38/00,C07H21/04,C07K14/47,C12N15/00,A61K37/02 CC			
<hr/>					
CC Strandedness:	Single;				
FH Key	Location/Qualifiers				
FT CDS	1..5412				
FEATURES	Location/Qualifiers				
source	/organism="unidentified"				
	#23244#				



180 .....TyrAspSerThrSerAspaspse 187  
 733 GCCCCTGCCAGATGCTGCTGCCGCTGGACCGCTGCTGCTGCCGCTAG 782  
 187 rAsnPheLeuAsnPro..... 192  
 783 TCGCAGCTCGCACCCCTCGCGGATCATGCAGACCCCTGGCCTCGCTGG 832  
 193 .....ProArgGly... 195  
 833 GCGCGAGGCGGCGGCCCTGCTCCCGGGACCCCTGGCCCCCGCGGGGAT 882  
 196 .....TrpAspHisThrAlaProGlyHisArg.....ThrPh 206  
 883 GTCTGCTTGAGAGATGCGGTGGTGGCTGAAAACTGCCTCACAG 932  
 206 eGluThrLysAspGlnProGluTyrAspSerThrAspGlyGluGlyAspT 223  
 933 CCTGACCCAGGCGGCGGGCACCGGCCACAGCGCGCTGGAAGCTGT 982  
 223 rpSerLeuTrpSerValCysSerValThrCysGlyAsnGlyAsnGlnLys 239  
 983 GGTCCCTGTGGGCGAATGCACGCGGACTCGGGGGAGGCCCTCCAGACG 1032  
 240 ArgThrArgSerCys.....GlyTyrAlaCysTh 249  
 1033 CGGAGCGCACCTGCTCCCGCGCGGCGTGGAGGCGGCGCTGCGA 1082  
 249 rAlaThr.....GluSerArgThrCysAspArgProAsnCysProGlyI 264  
 1083 GGGGTGCTGGAGGAGGTGCGCACTGCACCGCGAGGCCTGC..... 1125  
 264 leGluAspThrPheArgThrAlaAlaThrGluValSerLeuLeuAla... 279  
 1126 ..GGCCCCGCTGGGCGCACACGCTCCCGGAGCCAGTCCCTCGCGTCCACA 1173  
 280 .....GlySerGlu.....GluPh 284  
 1174 GATCCCGCGCGCGGAGAGCTGGGGACGAGCTGCAGCAGTTTGGGTT 1223  
 284 eAsnAlaThrLysLeuPheGluValAspThrAspSerCysGluArgTrpM 301  
 1224 CCCAGCCCCCAGCGGTGACCCAGCAGCGGAGGAGTGTCCTCCGCTGGA 1273  
 301 etSerCysLysSerGluPheLeuLysLysTyrMetHisLysValMetAsn 317  
 1274 GCGTGTGCTCCAGC..... 1287  
 318 AspLeuProSerCysProCysSerTyrProThrGluValAlaTyrSerTh 334  
 1288 .....ACCTGCGGC..... 1296  
 334 rAlaAspIlePheAspArgIleLysArgLysAspPheArgTrpLysAspA 351  
 1296 ..... 1296  
 351 laSerGlyProLysGluLysLeuGluIleTyrLysProThrAlaArgTyr 367  
 1297 .....GAGGGCTGCAGACCCCGCGCGCTTC 1323  
 368 CysIleArgSerMetLeuSerLeuGlu...SerThrThrLeuAlaAlaGl 383  
 1324 TCGGTGCTCTCTCTACAGACGAGTGCAGCGGACCCCTCGCGGAGCA 1373  
 383 nHisCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLysGlyA 400  
 1374 GCGGCTGTGC.....AACAACTCTCCGTGCGCCAGTGCATGGTG 1414  
 400 laGlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuHisTyr 416  
 1415 CCTGG.....GATGAGTGGTGC..... 1431  
 417 LysValAspValLeuProTrpIleIleCys.....LysGl 428

1432 .....CCTGGAGCCTCTGCTCCAGCACCTGTGGCGCGTGG 1466  
 428 yAspTrpSerArgTyrAsnGluAlaArgProProAsnAsnGlyGlnLysC 445  
 1467 CTTTCGGGATCGCACGCGCACCTGCAGGCCCCCCAGTTTGGGGGCAACC 1516  
 445 yThrGluSerProSerAspGlu 452  
 1517 CCTGTGAGGGCCCTGAGAGCAA 1539



/SDSI1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AA506014 +	128.00	174.77	0.25
/SDSI1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AA32053 +	128.00	173.71	0.25
/SDSI1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AA30310 +	128.00	173.71	0.25
/SDSI1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AA506013 +	128.00	171.81	0.25
/SDSI1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AA32014 +	128.00	165.80	0.65

```
seq_name: 708629.genecont/note_genecont/genecont Embry/MH2002.DNA.HM200207
seq_documentation_block:
ID   AAD23967 standard; cDNA; 1395 BP.
XX
XX
AC   AAD23967;
XX
XX
DT   26-MAR-2002 (first entry)
XX
XX   Human thrombospondin-like protein cDNA #1.
DE
DE
XX
KW   Human; thrombospondin-like protein; therapeutic; pharmacogenomic;
KW   drug screening; cosmetic; nutraceutical; anti-angiogenic; wound healing;
KW   endocrine; gene therapy; hyperthyroidism; hypothyroidism; ss.
KW
```

Accession	Species	Protein	Location/Qualifiers
OS	Homo sapiens.		
XX			
FH	Key		
FT	CDS		1..1395
FT			/*tag= a
FT			/product= "Thrombospondin-like protein"
FT			/transl_except= (pos:364..366, aa:Xaa)

FT misc\_feature 364 /tag= b  
 FT /note= "This degenerate base represents A-G  
 FT /note= "This degenerate base represents A-G  
 FT transition polymorphism"  
 FT 365  
 FT misc\_feature  
 FT /tag= c  
 FT /note= "This degenerate base represents A-T  
 FT transition polymorphism"  
 FT 535  
 FT misc\_feature  
 FT /tag= d  
 FT /note= "This degenerate base represents A-G  
 FT transition polymorphism"  
 XX  
 PN WO200190179-A2.  
 XX  
 XX 29-NOV-2001.  
 PD  
 XX  
 XX 23-MAY-2001; 2001WO-US16786.  
 XX  
 XX 23-MAY-2000; 2000US-206415P.  
 PR  
 XX (LEXI-) LEXICON GENETICS INC.  
 PA  
 XX  
 XX Turner CA, Hilbun E, Donoho G, Friedrich G, Abuin A, Zambrowicz B;  
 PI Sands AT;  
 PI  
 XX  
 XX WPI; 2002-089923/12.  
 DR P-PSDB; AAE14397.  
 DR  
 XX  
 XX  
 PT New isolated nucleic acid molecule that encodes human  
 PT thrombospondin-like protein, useful in therapeutic, diagnostic and  
 PT pharmacogenomic applications, in drug screening and clinical trial  
 PT monitoring -  
 XX  
 XX Claim 2; Page 35; 39pp; English.  
 XX  
 XX The invention relates to nucleic acid molecule encoding human  
 CC thrombospondin-like protein. The polynucleotide and polypeptide of  
 CC the invention are useful in therapeutic, diagnostic and pharmacogenomic  
 CC applications, in drug screening, clinical trial monitoring, and in  
 CC cosmetic or nutraceutical applications. The polynucleotide is useful for  
 CC identifying coding sequences and mapping a unique gene to a particular

CC chromosome, to screen libraries, isolate clones, prepare cloning and  
CC sequencing templates, as hybridisation probes for screening libraries  
CC and assessing gene expression patterns, in microarrays and other assay  
CC formats to screen collections of genetic material from patients who  
CC have a particular medical condition, to identify mutations associated  
CC with a particular disease, in a diagnostic or prognostic assay, as  
CC antisense molecules, and as part of ribozymes and/or triple helix  
CC sequences that are useful for gene regulation, for detecting mutant  
CC proteins or inappropriately expressed proteins for the diagnosis of  
CC disease and in the molecular mutagenesis/revolution of proteins that  
CC are at least partially encoded by the polynucleotide. The  
CC thrombospondin-like protein or its modified processed form can  
CC be used as therapeutics e.g. anti-angiogenic agents, to promote wound  
CC healing and regulate endocrine functions. The polynucleotide  
CC can also be used in gene therapy and for treating disorders like  
CC hyperthyroidism and hypothyroidism. The present sequence is a cDNA  
CC encoding human thrombospondin-like protein.  
XX

SQ Sequence 1395 BP; 345 A; 431 C; 384 G; 232 T; 3 other;

alignment\_scores:  
Quality: 2513.00 Length: 464  
Ratio: 5.439 Gaps: 0  
Percent Similarity: 99.569 Percent Identity: 99.569

alignment\_block:

US-09-863-824-2 x AAD23967 ..

Align seg 1/1 to: AAD23967 from: 1 to: 1395

1 MetValArgLeuAlaAlaGluLeuLeuLeuLeuGlyLeuLeuLeu 17  
|||||  
1 ATGTGGCCCTGGCGCGAGCTGCTGCTGCTGGGGCTGCTGCT 50  
17 uThrLeuHisIleThrValLeuArgGlySerGlyAlaAlaAspGlyProA 34  
|||||  
51 CACGCTGCACATCACCTGCTGCGCGGCTGGGAGCCGCGACGGGCGG 100  
34 spAlaAlaAlaGlyAsnAlaSerGlnAlaGlnLeuGlnAsnLeuAsn 50  
|||||  
101 ACGGCGCCGCGGCAACGCCAGCCAGCCAGCTGCAGATAACCTCAAC 150  
51 ValGlySerAspThrThrSerGluThrSerPheSerLeuSerLysGluAl 67  
|||||  
151 GTGGGAAGTGACACACATCAGAAACCAGCTTTTCTCTCTCAAGAAGC 200  
67 proArgGluHisLeuAspHisGlnAlaAlaHisGlnProPheProArgp 84  
|||||  
201 ACCAAGGAGGACATCTGGACCAACAGGCTGCACCAACCTTCCCCAGAC 250  
84 roArgPheArgGlnGluThrGlyHisProSerLeuGlnArgAspPhePro 100  
|||||  
251 CGCGATTCCGACAGAGACGGGCGACCTTCATTGCAAAAGAGATTCCCC 300  
101 ArgSerPheLeuLeuAspLeuProAsnPheProAspLeuSerLysAlaAs 117  
|||||  
301 AGATCTTCTCCTTGATCTACCAAACTTCCAGATCTTCCAAAGCTGA 350  
117 pIleAsnGlyGln\*\*\*ProAsnIleGlnValThrIleGluValValAspG 134  
|||||  
351 TATCAATGGCAGRWTCCTCAATATCCAGGTCCACATAGAGTGGTGCAGC 400  
134 lyProAspSerGluAlaAspLysAspGlnHisProGluAsnLysProSer 150  
|||||  
401 GTCCCTGACTCTGAAGCAGATAAAGATCAGCATCCGGAGAATAAGCCAGC 450  
151 TrpSerValProSerProAspTrpArgAlaTrpTrpGlnArgSerLeuSe 167  
|||||  
451 TGGTCAGTCCCATCCCTCCGACTGGCGGCGCTGGTGGCAGAGTCCCTGTC 500  
167 rLeuAlaArgAlaAsnSerGlyAspGlnAspTyr\*\*\*TyrAspSerThrs 184  
|||||

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT: AAD23969

seq\_documentation\_block:

501 CTTGGCCAGGGCAAAACAGCGGGACAGACTACRAGTACGACAGTACCT 550  
184 erAspAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAla 200  
|||||  
551 CAGACGACAGCAACTTCTCAACCCCCCAGGGGTGGGACCATACAGCC 600  
201 ProGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerTh 217  
|||||  
601 CCAGGCCACCGGACTTTTGAACCAAGATCAGCCAGAAATATGATTCCAC 650  
217 rAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerValThrCysG 234  
|||||  
651 AGATGCGGAGGTGACTGGAGTCTCTGGTCTGTCTGACGCGTACCTGG 700  
234 lyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAla 250  
|||||  
701 GGACCGCAACCAAGAACCGGACCGCTCTGTGGCTACGCGTGCACCTG 750  
251 ThrGluSerArgThrCysAspArgProAsnCysProGlyIleGluAspTh 267  
|||||  
751 ACAGAAATCGAGGACCTGTGACCGTCCAACTGCCAGGAATTGAAGACAC 800  
267 rPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerGluGluP 284  
|||||  
801 TTTTAGACAGCTGCCACCAAGTGTGCTGTGGGGAAGCGAGGACT 850  
284 heAsnAlaThrLysLeuPheGluValAspThrAspSerCysGluArgTrp 300  
|||||  
851 TTAATGCCACCAACTGTTTGAAGTTGACACAGACAGCTGTGAGCGCTGG 900  
301 MetSerCysLysSerGluPheLeuLysLysTyrMetHisLysValMetAs 317  
|||||  
901 ATGAGCTGCAAAAGCGAGTTCTTAAAGAGTACATGCACAAAGGTGATGA 950  
317 nAspLeuProSerCysProCysSerTyrProThrGluValAlaTyrSerT 334  
|||||  
951 TGACCTGCCAGCTGCCCTGCTCTACCCACTGAGGTGGCCTACAGCA 1000  
334 hrAlaAspIlePheAspArgIleLysArgLysAspPheArgTrpLysAsp 350  
|||||  
1001 CGCCCGACATCTTCGACCGCATCAAGCGCAAGGACTTCCCGTGGAGGAC 1050  
351 AlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAlaArgTy 367  
|||||  
1051 GCCAGCGGGCCCAAGGAGAGCTGGAGATCTACAAGCCCACTGCCCGGTA 1100  
367 rCysIleArgSerMetLeuSerLeuGluSerThrThrLeuAlaAlaGlnH 384  
|||||  
1101 CTGCATCCGCTCCATGCTGCTCCCTGGAGAGCACCGCTGGCGGCACAGC 1150  
384 isCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLysGlyAla 400  
|||||  
1151 ACTGCTGTACGGCGACACATGTCAGCTCATCACCAGGGGCAAGGGGCG 1200  
401 GlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuHisTyrLy 417  
|||||  
1201 GGCACGCCCAACCTCATCAGCACCGAGTTCTCCGCGGAGCTCCACTACAA 1250  
417 sValAspValLeuProTrpIleIleCysLysGlyAspTrpSerArgTyrA 434  
|||||  
1251 GGTGAGCTCCTGCCCTGGATTATCTCAAGGGTGTACTGGAGCAGGTATA 1300  
434 snGluAlaArgProProAsnAsnGlyGlnLysCysThrGluSerProSer 450  
|||||  
1301 ACGAGGCCCGGCTCCCAACACCGACAGAGTGCACAGAGAGCCCTCG 1350  
451 AspGluAspTyrIleLysGlnPheGlnGluAlaArgGluTyr 464  
|||||  
1351 GACGAGGACTACATCAAGCAGTTTCAAGAGGCCAGGGAATAT 1392



267 rPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerCluGluP 284  
|||||  
801 TTTTAGGACAGCTCCACCGAAGTGAGTCTGCTTGGCGGAAGCGAGGAGT 850  
284 heAsnAlaThrLysLeuPheGluValAlaThrThrAspSerCys 297  
|||||  
851 TTATGCCACCAACTGTTGAAGTGTGCTCCCGAGCATGT 891

seq\_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH99403

seq\_documentation\_block:  
ID AAH99403 standard; cDNA; 739 BP.  
AC AAH99403;  
DT 16-OCT-2001 (first entry)  
XX Human protein encoding cDNA sequence SEQ ID NO:238.  
XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiagregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;  
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder; SS.  
XX Homo sapiens.  
XX WO200153455-A2.  
XX 26-JUL-2001.  
XX 22-DEC-2000; 2000WO-US35017.  
XX 23-DEC-1999; 99US-0471275.  
XX 21-JAN-2000; 2000US-0488725.  
XX 25-APR-2000; 2000US-0552317.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Drmanac RT;  
XX WPI; 2001-457603/49.  
XX P-PSDB; AAM25462.  
XX Isolated human polynucleotides encoding polypeptides, useful for the  
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
XX Claim 1; Page 403-404; 1217pp; English.  
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
XX AAM25963. The proteins can have activities based on the tissues and  
XX cells they are expressed in, such as: antiinflammatory; antirheumatic;  
XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;  
XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
XX cardiovascular; antianaemic; antiagregant; haemostatic; vulnerary;  
XX antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;  
XX antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
XX encoding them can be used in gene therapy, antisense therapy and vaccine  
XX production. The proteins and polynucleotides are useful for screening for  
XX agonists or antagonists of a protein and for the treatment and diagnosis  
XX of disorders associated with the activity of a protein e.g. inflammation,  
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,

CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
CC anaemia, platelet disorders, thrombocytopaenia, wounds, ulcers, allergic  
CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
CC neurological disorders.  
XX  
SQ Sequence 739 BP; 187 A; 225 C; 206 G; 121 T; 0 other;

alignment\_scores:  
Quality: 1353.00 Length: 246  
Ratio: 5.522 Gaps: 0  
Percent similarity: 99.593 Percent identity: 99.593

alignment\_block:  
US-09-863-824-2 x AAH99403  
Align seg 1/1 to: AAH99403 from: 1 to: 739  
175 AspGlnAspTyr\*\*\*TyrAspSerThrSerAspAspSerAsnPheLeuAs 191  
|||||  
2 GACCAGGACTACAACTAGCAGCAGTACCTCAGCAGCAGCAGCAGTCCCTCAA 51  
191 nProProArgGlyTrpAspHisThrAlaProGlyHisArgThrPheGluT 208  
|||||  
52 CCCCCCAGGGGGTGGGACCATACAGCCCCAGGCCAGCGGACTTTTGGAAA 101  
208 hrLysAspGlnProGluTyrAspSerThrThrAspGlyGluGlyAspTyrSer 224  
|||||  
102 CCAAAGATCACGCCAGAAATATATGCCAGATGATGGAGGAGGAGTGAGT 151  
225 LeuTrpSerValCysSerValThrCysGlyAsnGlyAsnGlnLysArgTh 241  
|||||  
152 CTCTGGTCTGTCTGACGGCTCACCTCGGGGAACGGCAACCAAGAAACGGAC 201  
241 rArgSerCysGlyTyrAlaCysThrAlaThrGluSerArgThrCysAspA 258  
|||||  
202 CCGGTCTTGTGTGTACGGCTGCTGCAACAGAAATCGAGGACCTGTGACC 251  
258 rGProAsnCysProGlyIleGluAspThrPheArgThrAlaAlaThrGlu 274  
|||||  
252 GTCCAAACTGCCAGGAATTTGAAGACACTTTTAGGACAGTGCCACCGAA 301  
275 ValSerLeuLeuAlaGlySerGluGluPheAsnAlaThrLysLeuPheG 291  
|||||  
302 GTGAGTCTGTGCTTGGGGAAGCGAGAGTTAATGCCACCAAACTGTTTGA 351  
291 uValAspThrAspSerCysGluArgTyrMetSerCysLysSerGluPheL 308  
|||||  
352 AGTTGACACAGACAGCTGTGAGCGCTGGATGAGCTGCCAACAGCGAGTTCT 401  
308 euLysLysTyrMetHisLysValMetAsnAspLeuProSerCysProCys 324  
|||||  
402 TAAAGAAGTACATGCACAGGTGATGATGACCTGCCAGCTGCCCTGC 451  
325 SerTyrProThrGluValAlaTyrSerThrAlaAspIlePheAspArg 341  
|||||  
452 TCCTACCCCACTGAGGTGGCTTACAGCAGCGCGGACATCTTCGACCGCAT 501  
341 euLysArgLysAspPheArgTyrLysAspAlaSerGlyProLysGluLysL 358  
|||||  
502 CAAGCGCAAGGACTTCCGCTGGAAGAGACGCCAGCGGGGCCAAGGAGAAC 551  
358 euGluIleTyrLysProThrAlaArgTyrCysIleArgSerMetLeuSer 374  
|||||  
552 TGGAGATCTACAAGCCCACTGCCCGGTACTGCATCCGCTCCATGCTGCC 601  
375 LeuGluSerThrThrLeuAlaAlaGlnHisCysCysTyrGlyAspAsnMe 391  
|||||  
602 CTGGAGAGACCAACCGCTGGCGGCAGCAGCACTGCTGCTACGGCGCAACAT 651



AD AAF93820 standard; cDNA; 2981 BP.

AC AAF93820;

XX 23-MAY-2001 (first entry)

DT Human cDNA encoding a membrane or secretory protein clone PSEC0137.

DE Human; secretory protein; membrane protein; vaccine; gene therapy;

KW rheumatoid arthritis; diabetes; ss.

XX Homo sapiens.

XX EP1067182-A2.

XX 10-JAN-2001.

XX 07-JUL-2000; 2000EP-0114090.

XX 08-JUL-1999; 99JP-0194179.

XX 11-JAN-2000; 2000JP-0118775.

XX 02-MAY-2000; 2000JP-0183766.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;

PI WPI; 2001-093989/11.

XX P-PSDB; AAB88393.

DR Nucleic acids encoding secretory proteins/membrane proteins, useful in

XX gene therapy or as candidate target molecules in drug development -

PT Claim 1; SEQ ID 153; 609pp + CD ROM; English.

XX This invention relates to nucleic acid sequences AAF93744 - AAF93916

XX which encode human secretory or membrane proteins represented by

CC AAB88317 - AAB88419. Included in the invention are primers

CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the

CC cDNA sequences of the invention. The invention also includes methods for

CC the production of antibodies directed against the proteins, and cDNA

CC sequences, which can be used in vaccines. The polynucleotide sequences

CC can be used in gene therapy. The polynucleotide sequences and the

CC diagnosis of diseases associated with inappropriate secretory

CC protein/membrane protein expression. The nucleic acids and complementary

CC sequences may also be used as DNA probes in diagnostic assays

CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the

CC presence of similar nucleic acid sequences in samples. They may also be

CC used to study the expression and function of secretory proteins/membrane

CC polypeptides and their role in metabolism. The polypeptides may be used

CC as antigens in the production of antibodies against them and in assays to

CC identify modulators (agonists and antagonists) of expression and

CC activity. The antibodies and antagonists may also be used as therapeutic

CC agents to down regulate expression and activity. The antibodies may also

25 ArgGlySerGlyAlaAlaAspGlyProAspAlaAlaAlaGlyAsnAlase 41  
||||| : : : : : ||||| : : : : :  
357 AGAGGTACTCGTTGGCTGGAGCTGCAGAGCTGCCGGGATGGC... 404  
41 rGlnAlaGlnLeuGlnAsnLeuVal..... 51  
405 .....CAGCACAACTTGATACCCCTAACCTGATACCCAG 441  
52 .....GlySerAspThr 55  
442 GCTTACGCTCCAGATCCTAGGCCTCTGAGGGAAGAGGAGGACG 491  
56 ThrSerGluThrSerPheSerLeuSerLysGluAlaProArgGluHisLe 72  
||||| : : : : : : : : : : : : : : :  
492 ACTGCTCCCAACCCAGCCTGAGGAGGAGTACACACATGGATGT 541  
72 uAspHisGlnAlaAlaHisGlnProPheProArgProArgPheArgGlnG 89  
||||| : : : : : ||||| : : : : :  
542 GGACTGTACTGAGCCAGCAGCCCTGACCCAGG.....G 576  
89 lu.ThrGlyHisProSerLeuGlnArgAspPheProArgSerPheLeuLe 105  
|| : : : : : ||||| : : : : :  
577 AATGCCAGCCTCCAGGAGCCAGGAGGTACTCC.....TTGCTGCT 620  
105 uAspLeuProAsnProAspLeuSerLysAlaAspIleAsnGlyGln\* 122  
||||| : : : : : ||||| : : : : :  
621 GGAGCTCAGAAAGCTGCCAATTTGGTCCACGCAACCTTGATACCCCTA 670  
122 \*\*ProAsnIleGlnValThrIleGluValValAspGlyProAspSerGlu 138  
||||| : : : : : ||||| : : : : :  
671 ACCTGTATACCAAGTCAACATCAAGTGTGGAGGACCCAGGCGGAG 720  
139 AlaAspLysAsp.....GlnHisProGluAsnLysPro..... 149  
||||| : : : : : ||||| : : : : :  
721 GTGTGATAGACCTGTGGTGTAGCCCAATCCCGCCAGGATAC 770  
150 ....SerTrpSerValProSerProAspTTPArgAlaTTP..... 162  
||||| : : : : : ||||| : : : : :  
771 CCTTATGCTGG...CTGCCCGCCCTC...TGCCCTCTCTCTGGGGAGACT 814  
163 ....GlnArgSerLeuSerLeuAlaArgAlaAsnSerGly..... 174  
815 ACAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 864  
175 .....AspGlnAspTyr..... 178  
865 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 914  
179 .....\*\*\*TyrAspSerThrS 184  
915 GCACAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 964  
184 eAspAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAla 200  
: : : : : : : : : : : : : : :  
965 CAGAC.....AACTGGGACAGGAGGAGGAGGAGGAGGAGGAGGAGG 993  
201 ProGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerTh 217  
||||| : : : : : ||||| : : : : :  
994 CCGGGGAGTGGTCTTC.....AAGGATTCGTACGTACGAC...TA 1034  
217 rAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerValThrCysG 234  
: : : : : : : : : : : : : : :  
1035 TCAGCCTCAGAGGAGTGGAGTCCCTCTCCCTGAGTGGGAACTGCA 1084  
234 lYAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAla 250  
: : : : : : : : : : : : : : :  
1085 GCATGGCAAGCAGCAGGAGGAGTCCGCGCTGTGGCTATGGCTGCTGCC 1134  
251 ThrGluSerArgThrCysAspArgProAsnCysProGlyIleGlu..... 265  
||||| : : : : : ||||| : : : : :  
1135 ACCGAGACCCGTACCTGTGACCTGCCCTCTCTCTGCTGCTGCTGCTGCT 1184  
266 AspThrPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerG 282

alignment\_scores:  
Quality: 929.50 Length: 500  
Ratio: 2.923 Gaps: 18  
Percent Similarity: 63.600 Percent Identity: 41.600

alignment\_block:

US-09-863-824-2 x AAF93820 ..

Align seg 1/1 to: AAF93820 from: 1 to: 2981

```
|||||: : : ||| : : |||||||
1185 GGACACCTTGGCGCTCCCACTGAGGAGTGAAGCTCCTGGCCGCG.... 1230
282 luGluPheAsnAlaThrLysLeuPheGluValAspThrAspSerCysGlu 298
|||||: : : : : |||||||
1231 .....AATGCTACGAGCATGATGATCAAGATGTGGACAGCTGTGAG 1272
299 ArgTrpMetSerCysLysSerGluPheLeuLysLysTyrMetHisLysVa 315
|||||: : : : : |||||||
1273 AAGTGGCTGAACCTCAAGAGCGACTTCCTAATCAAGTATCTGAGCCAGAT 1322
315 lMetAsnAspLeuProSerCysProCysSerTyrProThrGluValAla 332
: : : : : |||||||
1323 GCTCGGGACCTGCCAGCTCCCGTGTGCTTACCCTAGTGAGGCCATGG 1372
332 yrSerThrAlaAspIlePheAspGlyLeuLysArgLysAspPheArgTrp 348
: : : : : |||||||
1373 ACAGCCCTGTGAGCTACAGACGAGCACACGAGCGCGCAGCTTCGGTGG 1422
349 LysAspAlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAl 365
: : : : : |||||||
1423 AGGATGCCAGTGGCCCTCGCGAGCGCTGGACATCTACCAGCCACGCG 1472
365 aArgTyrCysIleArgSerMetLeuSerLeuGluSerThrThrLeuAla 382
|||||: : : : : |||||||
1473 CGCGTCTGCTGCGTTCATGCTGCTGTGGGAGAGCAGCACACTGGCGG 1522
382 laGlnHisCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLys 398
|||||: : : : : |||||||
1523 CCCAGCACTGCTGTATGACGAGCAGCAGCGGCTGCTGACCCGTGGCAAG 1572
399 GlyAlaGlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuHi 415
|||||: : : : : |||||||
1573 GCGCGCGCATGCCCAACCTCATCAGCACCGCACTTCTCACCTAAGCTGA 1622
415 sTyrLysValAspValLeuProTrpIleIleCysLysGlyAspTrpSera 432
|||||: : : : : |||||||
1623 CTTCAAGTTCGACAGCAGCGCTGATCTGCTGCAAGGGGAGCTGGAGCC 1672
432 rgTyrAsnGluAlaArgProProAsnAsnGlyGlnLysCysThrGluSer 448
|||||: : : : : |||||||
1673 GCCTCCAGCGTGCCTCCTCCCAACACAGCGCGGCGCTGCACCGACAA 1722
449 ProSerAspGluAspTyrIleLysGlnPheGlnGluAlaArgGluTyr 464
|||||: : : : : |||||||
1723 CCCCTGGAGGAGGAGTACCTAGCACAGTTCGAGGAGGCCAAGGAGTAC 1770
```

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF29348

seq\_documentation\_block:

ID AAF29348 standard; cDNA; 2981 BP.

XX

AC AAF29348;

XX

XX 20-APR-2001 (first entry)

XX Human proliferation differentiation factor cDNA sequence.

XX Human; proliferation differentiation factor; haematopoietic function; ss.

XX Homo sapiens.

OS

XX WO200104312-A1.

XX

XX 18-JAN-2001.

PD

XX 06-JUL-2000; 2000WO-JP04514.

XX

XX 08-JUL-1999; 99JP-0194179.

PR

XX 18-OCT-1999; 99US-0159586.

XX

XX (HELI-) HELIX RES INST.

PA

XX

PI Ota T, Isogai T, Nishikawa T, Kawai Y, Yoshida K, Masuho Y;  
XX WPI: 2001-138354/14.  
DR P-PSDB; AAB49765.  
XX  
PT Proliferation-differentiation factor protein encoded by PSEC137 cloned  
PT from human cDNA library, being hematopoietic factor for inducing  
PT differentiation of blood cells, used to maintain hematopoietic function  
XX  
XX  
PS Claim 1; Page 36-41; 49pp; Japanese.  
XX  
CC This invention relates to polynucleotide sequence PSEC137, which encodes  
CC a human proliferation differentiation factor protein. Included in the  
CC invention is a vector containing the cDNA sequence, a transformant  
CC containing the vector, and a process for producing the protein. The  
CC protein together with its encoded DNA and antisense DNA are used in drugs  
CC to maintain the haematopoietic function. The present sequence represents  
CC cDNA encoding the proliferation differentiation factor protein.  
XX  
SQ Sequence 2981 BP; 609 A; 925 C; 887 G; 560 T; 0 other;

alignment\_scores:

Quality: 929.50 Length: 500  
Ratio: 2.923 Gaps: 18

Percent Similarity: 63.600 Percent Identity: 41.600

alignment\_block:

US-09-863-824-2 x AAF29348 ..

Align seg 1/1 to: AAF29348 from: 1 to: 2981

```
25 ArgGlySerGlyAlaAlaAspGlyProAspAlaAlaGlyAsnAlaSe 41
||||| : : : : ||| ||||||| : :
357 AGAGGTTACTCCGTTCGGTGGGAGCTGCAGAAAGCTCGGGGATGGC... 404
41 rGlnAlaGlnLeuGlnAsnAsnLeuAsnVal..... 51
||||| : : : : |||||||
405 .....CAGCACACCTTGAGTACCCCTAACCCCTGATACCCAG 441
52 .....GlySerAspThr 55
||||| : : : : |||
442 GCTTCAGCCTCCCGATCTCTAGGCCTCTGAGGGAAGAGGAGGACG 491
56 ThrSerGluThrSerPheSerLeuSerLysGluAlaProArgGluHisLe 72
||||| : : : : |||
492 ACTGCTCCCCAGAACCCACCTGCAGGCAGAGCTACACCAACATGGATGT 541
72 uAspHisGlnAlaAlaHisGlnProPheProArgProArgPheArgGlnG 89
||||| : : : : |||||||
542 GGACTGTCACCTGAGCCAGCAGCCCTTGACCCCGG.....G 576
89 lu.ThrGlyHisProSerLeuGlnArgAspPheProArgSerPheLeuLe 105
|| : : : : |||
577 AATGCCAGCCTCCCGAGGACCCAGGAGGTACTCC.....TTGCTGCT 620
105 uAspLeuProAsnPheProAspLeuSerLysAlaAspIleAsnGlyGln* 122
||||| : : : : |||
621 GGAGCTGCAGAAAGTGCAGAAATTGGTCCACGCAACCTTGAGTACCCCTA 670
122 **ProAsnIleGlnValThrIleGluValValAspGlyProAspSerGlu 138
||||| : : : : |||||||
671 ACCCTGATAACCAAGTGACCATCAAGGTGTGGAGGACCCCGCCGCGAG 720
139 AlaAspLysAsp.....GlnHisProGluAsnLysPro..... 149
: : : : : |||
721 GTGTGATAGACCTGTTGGCTGAGCCAGCAATCCCCCGCCCGCAGGATAC 770
150 ....SerTrpSerValProSerProAspTrpArgAlaTrpTrp..... 162
||||| : : : : |||
771 CTTTAGCTGG...CTGCCCGCCCTC...TGGCCCTTCTCTCTGGGAGACT 814
```



163 .....GlnArgSerLeuSerLeuAlaArgAlaAsnSerGly..... 174  
 815 ACAAGAGAGAAAAGACAGAGCCGCCAGGAGAGAGGGGAGAAAAG 864  
 175 .....AspGlnAspTyr..... 178  
 865 GAGGAAGACGAGGACTATCTTCAGAGATATCGAGGGTGAGGATCAAGA 914  
 179 .....\*\*\*TyrAspSerThrS 184  
 915 GGACAAAGAGAGAGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 964  
 184 erAspAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAla 200  
 965 CAGC.....AAGTGGGACAGGGCTG.....CTGGCC 993  
 201 ProGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerTh 217  
 994 CCGGGGATGGGTCTTC.....AAGGATTCGTGCTACGTACGAC...TA 1034  
 217 rAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerValThrCysG 234  
 1035 TGAGCCTCAGAGAGAGTGGAGTCCCTGCTCCCTGCAGTGGGAAGTCA 1084  
 234 lAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAla 250  
 1085 GCATGGCAAGCAGCAGAGAGTCCGCCCTGTGGCTATGGCTGCACTGCC 1134  
 251 ThrGluSerArgThrCysAspArgProAsnCysProGlyIleGlu..... 265  
 1135 ACCGAGACCCGTACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1184  
 266 AspThrPheArgThrAlaAlaThrGluValSerLeuAlaGlySerG 282  
 1185 GGACACCTTGGGCCCTCCCGAGTGGAGAGTGGAGTCTCTGGCCCGC... 1230  
 282 lGluPheAsnAlaThrLysLeuPheGluValAspThrAspSerCysGlu 298  
 1231 .....AATGCTACGGACATGCATCAAGATGTGGACAGCTGTGAG 1272  
 299 ArgTrpMetSerCysLysSerGluPheLeuLysLysTyrMetHisLysVa 315  
 1273 AGTGCTGAAGTGCAGAGAGAGTCTCTCAAGTATCTGAGCCAGAT 1322  
 315 lMetAsnAspLeuProSerCysProCysSerTyrProThrGluValAla 332  
 1323 GCTGGGGACCTGCCAGCTGCCGTGCTGCTGCTGCTGCTGCTGCTGCTG 1372  
 332 yrSerThrAlaAspPheAspArgIleLysArgLysAspPheArgTrp 348  
 1373 ACAGCCCTGTGAGCCTACAGGAGAGCAGCAGGCGCGAGCTTCGGTGG 1422  
 349 LysAspAlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAl 365  
 1423 AGGATGCCAGTGGCCCTCGGAGCGCTGGACATCTACCGCCAGCCGCG 1472  
 365 aArgTyrCysIleArgSerMetLeuSerLeuGluSerThrThrLeuAla 382  
 1473 GCCTTCTGCCTGCTTCCATGCTGTGCGGAGAGCAGCAGTGGCGG 1522  
 382 lGlnHisCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLys 398  
 1523 CCAGCAGCTGCTGCTATGAGGAGAGCAGCGGCTGTGCTACCGTGGCAG 1572  
 399 GlyAlaGlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuH 415  
 1573 GGGCGCGGATGCCCAACCTCATCAGCAGCAGCTTCTCACCCTAAGTGCA 1622  
 415 sTyrLysValAspValLeuProTrpIleLysLysGlyAspTrpSerA 432  
 1623 CTTCAAGTTGAGACAGCAGCCCTGGATCTGTGCAAGGGGGAGCTGGAGCC 1672  
 432 rgTyrAsnGluAlaArgProProAsnAsnGlyGlnLysCysThrGluSer 448

1673 GCCTCCACGCTGTGCTCCCTCCCAACGCGGAGCTGCACCGCAAC 1722  
 449 ProSerAspGluAspTyrIleLysGlnPheGlnGluAlaArgGluTyr 464  
 1723 CCCTGGAGGAGGAGTACCTAGCACAGTTCAGGAGGCCAAGGAGTAC 1770  
 seq\_name: /SIDS1/gcdata/hold-geneseq/geneseq-emb1/NA2002.DAT.AAD23368  
 seq\_documentation\_block:  
 ID\_ AAD23968 standard; cDNA; 495 BP.  
 XX  
 AC AAD23968;  
 DT 26-MAR-2002 (first entry)  
 XX Human thrombospondin-like protein cDNA #2.  
 DE  
 XX Human; thrombospondin-like protein; therapeutic; pharmacogenomic;  
 KW drug screening; cosmetic; nutraceutical; anti-angiogenic; wound healing;  
 KW endocrine; gene therapy; hyperthyroidism; hypothyroidism; ss.  
 XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..495  
 FT /\*tag= a  
 FT /product= "Thrombospondin-like protein"  
 XX  
 PN WO200190179-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 23-MAY-2001; 2001WO-US16786.  
 XX  
 PR 23-MAY-2000; 2000US-206415P.  
 XX  
 PA (LEXI-) LEXICON GENETICS INC.  
 XX  
 XX Turner CA, Hilbun E, Donoho G, Friedrich G, Abuin A, Zambrowicz B;  
 PI Sands AT;  
 XX  
 DR WPI: 2002-089923/12.  
 DR P-PSDB; AAE14398.  
 DR  
 XX New isolated nucleic acid molecule that encodes human  
 PT thrombospondin-like protein, useful in therapeutic, diagnostic and  
 PT pharmacogenomic applications, in drug screening and clinical trial  
 PT monitoring -  
 XX  
 PS Claim 1; Page 37; 39pp; English.  
 XX  
 CC The invention relates to nucleic acid molecule encoding human  
 CC thrombospondin-like protein. The polynucleotide and polypeptide of  
 CC the invention are useful in therapeutic, diagnostic and pharmacogenomic  
 CC applications, in drug screening, clinical trial monitoring, and in  
 CC cosmetic or nutraceutical applications. The polynucleotide is useful for  
 CC identifying coding sequences and mapping a unique gene to a particular  
 CC chromosome, to screen libraries, isolate clones, prepare cloning and  
 CC sequencing templates, as hybridisation probes for screening libraries  
 CC and assessing gene expression patterns, in microarrays and other assay  
 CC formats to screen collections of genetic material from patients who  
 CC have a particular medical condition, to identify mutations associated  
 CC with a particular disease, in a diagnostic or prognostic assay, as  
 CC antisense molecules, and as part of ribozymes and/or triple helix  
 CC sequences that are useful for gene regulation, for detecting mutant  
 CC proteins or inappropriately expressed proteins for the diagnosis of  
 CC disease and in the molecular mutagenesis/evolution of proteins that  
 CC are at least partially encoded by the polynucleotide. The  
 CC thrombospondin-like protein or its modified processed form can  
 CC be used as therapeutics e.g. anti-angiogenic agents, to promote wound  
 CC healing and regulate endocrine functions. The polynucleotide  
 CC can also be used in gene therapy and for treating disorders like



CC hyperthyroidism and hypothyroidism. The present sequence is a cDNA  
 CC encoding human thrombospondin-like protein.  
 XX  
 SQ Sequence 495 BP; 126 A; 156 C; 139 G; 74 T; 0 other;

alignment\_scores:  
 Quality: 900.00 Length: 164  
 Ratio: 5.488 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-863-824-2 x AAD23968 ..

Align seg 1/1 to: AAD23968 from: 1 to: 495

301 MetSerCysLysSerGluPheLeuLysTyrMetHisLysValMetAs 317  
 |||||  
 1 ATGAGCTGCAAAAGCGAGTCTTAAGAAAGTACATGCACAAGGTGATGAA 50  
 317 nAspLeuProSerCysProCysSerTyrProThrGluValAlaTyrSert 334  
 |||||  
 51 TGACCTGGCCAGCTGCCCTCTCTACCCCACTGAGGTGGCTACAGCA 100  
 334 hrAlaAspIlePheAspArgIleLysArgLysAspPheArgTrpLysasp 350  
 |||||  
 101 CGGCGGACATCTTCGACGCATCAAGCGCAAGGACTTCGCTGGAAGGAC 150  
 351 AlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAlaArgTy 367  
 |||||  
 151 GCCAGCGGCGCAAGGAGAGATGGAGATCTACAAGGCCACTGCCCGGTA 200  
 367 rCysIleArgSerMetLeuSerLeuGluSerThrThrLeuAlaAlaGlnH 384  
 |||||  
 201 CTGATCCGCTCCATGCTGCTGAGAGACCAACGCTGGCGGCACAGC 250  
 384 IsCysCysTyrGlyAspAsnMetGlnLeuIleThrArgGlyLysGlyAla 400  
 |||||  
 251 ACTGCTGTACGGGACAAACATGCAGCTCATCACAGGGGCAAGGGGCG 300  
 401 GlyThrProAsnLeuIleSerThrGluPheSerAlaGluLeuHisTyrLy 417  
 |||||  
 301 GGCAGCGCCCAACCTCATCAGCACCGAGTCTCCGCGGAGCTCCACTACA 350  
 417 sValAspValLeuProTTPileIleCysLysGlyAspTrpSerArgTyrA 434  
 |||||  
 351 GGTGGACGCTCTGCGCTGGATATCTGCAAGGGTGACTGGAGCAGGTATA 400  
 434 snGluAlaArgProProAsnAsnGlyGlnLysCysThrGluSerProSer 450  
 |||||  
 401 ACGAGGCCGCGCTCCCAACAAGGACAGAAAGTGACAGAGAGCCCTCG 450  
 451 AspGluAspTyrIleLysGlnPheGlnGluAlaArgGluTyr 464  
 |||||  
 451 GACGAGGACTACATCAAGCAGTTCACAGAGGCCAGGGAATAT 492

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NA2001B.DAT.AAS70533

seq\_documentation\_block:

ID AAS70533 standard; cDNA; 448 BP.

XX  
 AC AAS70533;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE DNA encoding novel human diagnostic protein #6337.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.

XX 11-OCT-2001.  
 PD  
 XX 30-MAR-2001; 2001WO-US08631.  
 PF  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Drmanac RT, Liu C, Tang YT;  
 PI  
 XX WPI; 2001-639362/73.  
 DR P-PSDB; ABG06346.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 1; SEQ ID No 6337; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 448 BP; 123 A; 122 C; 123 G; 80 T; 0 other;

alignment\_scores:  
 Quality: 836.00 Length: 149  
 Ratio: 5.649 Gaps: 0  
 Percent Similarity: 99.329 Percent Identity: 99.329

alignment\_block:  
 US-09-863-824-2 x AAS70533 ..

Align seg 1/1 to: AAS70533 from: 1 to: 448

175 AspGlnAspTyr\*\*\*TyrAspSerThrSerAspSerAsnPheLeuAs 191  
 |||||  
 2 GACCAGGACTACAAGTACGACAGTACCTCAGACGACAGCAACTTCCTCAA 51  
 191 nProProArgGlyTrpAspHisThrAlaProGlyHisArgThrPheGluT 208  
 |||||  
 52 CCCCCCAGGGGTGGGACCACATACAGCCCCAGGCCCGGACTTTTGAA 101  
 208 hrLysAspGlnProGluTyrAspSerThrAspGlyGluGlyAspTrpSer 224  
 |||||  
 102 CCAAGATCAGCCAGAAATATGATCCACAGATGGCGAGGTCACGTGAGT 151  
 225 LeuTrpSerValCysSerValThrCysGlyAsnGlyAsnGlnLysArgTh 241  
 |||||  
 152 CTCTGGTCTGTCTGCGAGCTCACCTGCGGGAACGCAACACGAAACGGAC 201

241 rArgSerCysGlyTyrAlaCysThrAlaThrGluSerArgThrCysAspA 258  
 202 CGGTCCTTGGCTACGGCTGACCTGCAACAGATCGAGGACCTGTGACC 251  
 258 rGProAsnCysProGlyGlyLeuGluAspThrPheArgThrAlaAlaThrGlu 274  
 252 GTCCAAACTGCCAGCAATTAAGACACACTTTTAGGACAGCTGCCACCGAA 301  
 275 ValSerLeuLeuAlaGlySerGluGluPheAsnAlaThrLysLeuPheG1 291  
 302 GGTGACTGCTTCGGGGAAGGAGGAGTTTAAATGCCACCAACTGTTTGA 351  
 291 uValAspThrAspSerCysGluArgTrpMetSerCysLysSerGluPheL 308  
 352 AGTTGCACACAGACAGCTGTGAGCGCTGATGAGCTGCAAAAGCGAGTTCT 401  
 308 euLysLysTyrMethiLysValMetAsnAspLeuProSerCysPro 323  
 402 TAAAGAAGTACATGCACAAGGTGATGATGACCTGCCCGAGCTGCCCC 448

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAC76131

seq\_documentation\_block:  
 ID AAC76131 standard; cdna; 790 BP.

XX AAC76131;  
 AC  
 DT 08-FEB-2001 (first entry)  
 XX

DE Human ORFX ORF1686 polynucleotide sequence SEQ ID NO:3371.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.

XX Homo sapiens.  
 OS  
 XX WO200058473-A2.  
 PN

XX 05-OCT-2000.  
 PD  
 XX 31-MAR-2000; 2000WO-US08621.  
 PF  
 XX 31-MAR-1999; 99US-0127607.  
 PR  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 PR

XX (CURA-) CURAGEN CORP.  
 PA  
 XX Shinkets RA, Leach M;  
 PI  
 XX WPI; 2000-602362/57.  
 DR  
 DR P-PSDB; AAB41922.  
 DR

XX Novel nucleic acids and peptides derived from open reading frame x,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX Claim 5; Page 2553; 5507pp; English.  
 PS  
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antithyroid; and antianemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, diabetes mellitus,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 790 BP; 181 A; 238 C; 212 G; 157 T; 2 other;

alignment\_scores:  
 Quality: 440.50 Length: 91  
 Ratio: 5.122 Gaps: 1  
 Percent Similarity: 94.505 Percent Identity: 91.209

alignment\_block:  
 US-09-863-824-2 x AAC76131/rev ..

Align seg 1/1 to reverse of: AAC76131 from: 1 to: 790  
 160 AlaTrpTrpGlnArg.....SerLeuSerLeuAlaArgAlaAsnSe 173  
 275 GCCTGGTGGTCCAGATGCTCCCTTGGTGTCTTTGGCAGGGCAAA.AG 227  
 173 rGlyAspGlnAspTyr\*\*\*TyrAspSerThrSerAspAspSerAsnPhel 190  
 226 CGGGACACAGACATACAGTACGGCAGTACCTACAGACAGACAGCACTCC 177  
 190 euAsnProProArgGlyTrpAspHisThrAlaProGlyHisArgThrPhe 206  
 176 TCACCCCCCAGGGGTGGGACCATACAGCCCGCCAGCCACCGGACTTT 127  
 207 GluThrLysAspGlnProGluTyrAspSerThrAspGlyGluGlyAspTr 223  
 126 GAAACCAAGATCAGCCAGAAATATGATTCACAGATGGCGAGGGTGACTG 77  
 223 pSerLeuTrpSerValCysSerValThrCysGlyAsnGlyAsnGlnLysA 240  
 76 GAGTCTCTGGTCTGTCTGACGGTACCTCGGGGAACGGCAACCAAGAAC 27  
 240 rGThrArgSerCysGlyTyrAla 247  
 26 GGACCCGGTCTGTGGCTACGGC 4

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS34028

seq\_documentation\_block:  
 ID AAS34028 standard; cdna; 478 BP.

XX AAS34028;  
 AC  
 DT 17-DEC-2001 (first entry)  
 XX

XX Human cdna encoding a novel foetal antigen, SEQ ID No 552.  
 DE  
 DE Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;  
 KW immunomodulator; cardiovascular; cytostatic; nephroprotective;  
 KW cardiovascular; autoimmune disease; rheumatoid arthritis;  
 KW hyperproliferative disorder; breast neoplasm; cancer;  
 KW

KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; food additive.

KW	cardiovascular disorder; cardiac arrest; cerebrovascular disorder;	21-SEP-2000;	2000US-0234223
KW	cerebral ischaemia; angiogenesis; nervous system disorder;	21-SEP-2000;	2000US-0234274
KW	Alzheimer's disease; infection; ocular disorder; corneal infection;	25-SEP-2000;	2000US-0234997
KW	wound healing; epithelial cell proliferation; food additive.	25-SEP-2000;	2000US-0234998
XX		26-SEP-2000;	2000US-0235484
XX		27-SEP-2000;	2000US-0235834
OS	Homo sapiens.	27-SEP-2000;	2000US-0235836
XX		29-SEP-2000;	2000US-0236327
PN	WO200155312-A2.	29-SEP-2000;	2000US-0236367
XX		29-SEP-2000;	2000US-0236368
PD		29-SEP-2000;	2000US-0236369
XX		29-SEP-2000;	2000US-0236370
XX		02-OCT-2000;	2000US-0236802
PF	17-JAN-2001; 2001WO-US01321.	02-OCT-2000;	2000US-0237037
XX		02-OCT-2000;	2000US-0237038
PR	31-JAN-2000; 2000US-0179065.	02-OCT-2000;	2000US-0237039
PR	04-FEB-2000; 2000US-0180628.	02-OCT-2000;	2000US-0237040
PR	24-FEB-2000; 2000US-0184664.	02-OCT-2000;	2000US-0237040
PR	02-MAR-2000; 2000US-0186350.	02-OCT-2000;	2000US-0239935
PR	16-MAR-2000; 2000US-0189874.	13-OCT-2000;	2000US-0239935
PR	17-MAR-2000; 2000US-0190076.	13-OCT-2000;	2000US-0239937
PR	18-APR-2000; 2000US-0198123.	20-OCT-2000;	2000US-0240960
PR	19-MAY-2000; 2000US-0205515.	20-OCT-2000;	2000US-0241221
PR	07-JUN-2000; 2000US-0209467.	20-OCT-2000;	2000US-0241785
PR	28-JUN-2000; 2000US-0214886.	20-OCT-2000;	2000US-0241786
PR	30-JUN-2000; 2000US-0215135.	20-OCT-2000;	2000US-0241787
PR	07-JUL-2000; 2000US-0216647.	20-OCT-2000;	2000US-0241808
PR	07-JUL-2000; 2000US-0216880.	20-OCT-2000;	2000US-0241809
PR	11-JUL-2000; 2000US-0217487.	20-OCT-2000;	2000US-0241826
PR	11-JUL-2000; 2000US-0217496.	01-NOV-2000;	2000US-0244617
PR	14-JUL-2000; 2000US-0218290.	08-NOV-2000;	2000US-0246474
PR	26-JUL-2000; 2000US-0220953.	08-NOV-2000;	2000US-0246475
PR	26-JUL-2000; 2000US-0220964.	08-NOV-2000;	2000US-0246524
PR	14-AUG-2000; 2000US-0225118.	08-NOV-2000;	2000US-0246525
PR	14-AUG-2000; 2000US-0225119.	08-NOV-2000;	2000US-0246526
PR	14-AUG-2000; 2000US-0225268.	08-NOV-2000;	2000US-0246527
PR	14-AUG-2000; 2000US-0225268.	08-NOV-2000;	2000US-0246528
PR	14-AUG-2000; 2000US-0225447.	08-NOV-2000;	2000US-0246532
PR	14-AUG-2000; 2000US-0225757.	08-NOV-2000;	2000US-0246532
PR	14-AUG-2000; 2000US-0225758.	08-NOV-2000;	2000US-0246609
PR	14-AUG-2000; 2000US-0225759.	08-NOV-2000;	2000US-0246610
PR	18-AUG-2000; 2000US-0226279.	08-NOV-2000;	2000US-0246611
PR	22-AUG-2000; 2000US-0226681.	17-NOV-2000;	2000US-0249207
PR	22-AUG-2000; 2000US-0226868.	17-NOV-2000;	2000US-0249208
PR	22-AUG-2000; 2000US-0227182.	17-NOV-2000;	2000US-0249209
PR	23-AUG-2000; 2000US-0227009.	17-NOV-2000;	2000US-0249210
PR	30-AUG-2000; 2000US-0228924.	17-NOV-2000;	2000US-0249211
PR	01-SEP-2000; 2000US-0229287.	17-NOV-2000;	2000US-0249212
PR	01-SEP-2000; 2000US-0229343.	17-NOV-2000;	2000US-0249213
PR	01-SEP-2000; 2000US-0229344.	17-NOV-2000;	2000US-0249214
PR	01-SEP-2000; 2000US-0229345.	17-NOV-2000;	2000US-0249215
PR	05-SEP-2000; 2000US-0229509.	17-NOV-2000;	2000US-0249216
PR	05-SEP-2000; 2000US-0229513.	17-NOV-2000;	2000US-0249217
PR	06-SEP-2000; 2000US-0230437.	17-NOV-2000;	2000US-0249218
PR	06-SEP-2000; 2000US-0230438.	17-NOV-2000;	2000US-0249244
PR	08-SEP-2000; 2000US-0231242.	17-NOV-2000;	2000US-0249245
PR	08-SEP-2000; 2000US-0231243.	17-NOV-2000;	2000US-0249264
PR	08-SEP-2000; 2000US-0231244.	17-NOV-2000;	2000US-0249265
PR	08-SEP-2000; 2000US-0231413.	17-NOV-2000;	2000US-0249297
PR	08-SEP-2000; 2000US-0231414.	17-NOV-2000;	2000US-0249299
PR	08-SEP-2000; 2000US-0232080.	17-NOV-2000;	2000US-0251988
PR	08-		







```
1011 AAGCTATCTGTGGCTCTCTCTGTGATGAACATATCCAGCATGGTCTCTGGAA 1060
64  exLysGluAlaProArgGluHisLeuAspHisGlnAlaAlaHisGlnPro 80
   :: |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1061 CT...GAAGGGCTCGGCACCATCTGTGACCACTCTGCAGGACGATCCG 1107
81  PheProArgProArgPheArgGln.....GluThrGlyHis 92
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1108 AAAAGTGAAGGAAAGAACAGAGAGCTGTGTCAGTGAAGCGGCGCTC 1157
92  sProSerLeuGlnArg.....AspP 99
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1158 CCTCTGCTTTTACACATGGATCCAGTCCAGTACAAAGAACAGGAGTGGACT 1207
99  heProArgSerPheLeuLeuAspLeuPro.AsnPheProAspLeuSerLys 115
   :: |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1208 GTAGACAGTTCACAGAGTGTCTGTCAGGACCACTCGTTTACCATCTGCAA 1257
115  sAlaAspLeuAsnGlyGln***ProAsnIleGlnValThrIleGluValV 132
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1258 AAAGGTGTCTCTGTCATCATGCTCTGCTCCAAACGCCACAGTT..... 1300
132  aLAspGlyProAspSerGluAlaAspLysAspGlnHisProGluAsnLys 148
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1301 .....CCTGATGGTGAATGC.....TGCCCGACGGTCTGG 1330
149  ProSerTrpSerValProSerProAspTrpArgAlaTrpTrpGlnArgSe 165
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1331 CCCACGACTGTCT...GAGATGGCTGGTCTCCTCTGCTGAGTGGAC 1377
165  rLeuSerLeuAlaArgAlaAsnSerGlyAspGlnAsp.....Tyr***T 180
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1378 CTCCTGCTCTGCCACATGTGGCAATGGAATTCCAGCAACGTGCTCTCTCT 1427
180  yrAspSerThrSerAsp..... 185
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1428 CTGACAGCTCTCAACAAACAGATGCGAGGCTCTTCCGTATACAGACGAGACC 1477
186  .....AspSerAsnPheLeuAsnProProArgGlyTr 196
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1478 TGCCACATTCAGGAGTGTGACAAAGATTT...AAACAGGATGGTGGCTG 1524
196  pAspHisThrAlaProGlyHisArgThrPheGluThr..... 208
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1525 GAGTCACTGGTCTCCATGCTGTCTGTTCTGTGACCTGTGTGACGGTG 1574
209  .....LysAspGlnProGluTrpAspSer 216
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1575 TGATCACAAGGATCGTCTCTGCAACTCCCGGAGACCAAGCCTGCAGAAAGAGCG 1624
217  ThrAspGlyGluGlyAsp..... 222
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1625 AAGCCCTGTGAAGGTGAAGCGCGGAGACCAAGCCTGCAGAAAGAGCG 1674
223  .....TrpSerLeuTrpSerValCys 230
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1675 CTGCCCAATTAATGAGGCTGGGGTCCCTGTGTCACCATGGGACATCTGCT 1724
230  erValThrCysGlyAsnGlyAsnGlnLysArgThrArgSerCys..... 244
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1725 CTGTACCTGTGGAGGAGGTGCAGAGACGCGCGACTCTGTGAACAC 1774
245  .....GlyTyrAlaCysThr.....AlaThrGluSe 253
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1775 CCCACACCCAGTTGGAGGCAAGACGTGTGTGGCGATGTGACAGAAAA 1824
253  rArgThrCysAspArgProAsnCysPro 262
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1825 TCAAGTTTGAACAACAGAGGACTGCCCCA 1852
```

seq\_name: /SIDSL/gcdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.AAC90311

seq\_documentation\_block:

```
ID AAC90311 standard; DNA; 4339 BP.
XX
AC AAC90311;
XX
DT 19-MAR-2001 (first entry)
XX
XX M87276 cDNA clone.
XX
XX METH; metalloprotease; thrombospondin; angiogenesis inhibition;
KW cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;
KW plaque neovascularisation; telangiectasia; haemophilic joint; ESR;
KW angiofibroma; fibromuscular dysplasia; expressed sequence tag;
KW Crohn's disease; atherosclerosis; birth control; ss.
XX
OS Unidentified.
XX
XX WO200071577-A1.
XX
XX 30-NOV-2000.
XX
XX 25-MAY-2000; 2000WO-US14462.
XX
XX 25-MAY-1999; 99US-0318208.
XX
XX 20-JUL-1999; 99US-0144882.
XX
XX 10-AUG-1999; 99US-0147823.
XX
XX 13-AUG-1999; 99US-0373658.
XX
XX 22-DEC-1999; 99US-0171503.
XX
XX 22-FEB-2000; 2000US-0183792.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX (IRUE/) IRUELA-ARISPE L.
XX
XX (HAST/) HASTINGS G A.
XX
XX (RUBE/) RUBEN S M.
XX
XX (JONA/) JONAK Z L.
XX
XX (TRUL/) TRULLI S H.
XX
XX (FORN/) FORNWALD J A.
XX
XX (TERR/) TERRETT J A.
XX
XX IrueLa-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
XX Fornwald JA, Terrett JA;
XX
XX WPI; 2001-025136/03.
XX
XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
XX inhibit angiogenesis in the treatment of disorders such as cancer,
XX rheumatoid arthritis and psoriasis -
XX
XX Claim 14; Pages 720-723; 768pp; English.
XX
XX The present invention relates to human METH1 and METH2, (ME for
XX metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
XX The present sequence is an expressed sequence tag (EST) for METH. METH
XX can be used for inhibiting angiogenesis in an individual, and for
XX treating cancer, benign tumours, an ocular angiogenic disease,
XX rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
XX vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
XX scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
XX coronary collaterals, cerebral collaterals, arteriovenous malformations,
XX ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
XX neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
XX fibromuscular dysplasia, wound granulation, Crohn's disease or
XX atherosclerosis. METH can also be used in birth control. METH can also
XX be used in diagnostic methods for the prognosis of cancer.
XX
XX Sequence 4339 BP; 1109 A; 1201 C; 1144 G; 885 T; 0 other;
```

[illegible]

```

1625 AACCCCTGTGAAGGTGAAGCCCGGAGACCAAAAGCCTGCAAGAAAGACGC 1674
223 .....TipSerLeuTrpSerValCysS 230
1675 CTGCCCAATTAATGAGGCTGGGGTCCTGTCACCATGGACATCTGCT 1724
230 erValThrCysGlyAsnGlyAsnGlnLysArgThrArgSerCys..... 244
1725 CTGTCACTGTGTGAGGAGGATGGAGAGACGACGCCGACTCTGTATCAAC 1774
245 .....GlyTyralaCysThr.....AlaThrGlucose 253
1775 CCACACCCCGCTTTGGAGGCAAGACTGTCTTGGCGATGTGACAGAAA 1824
253 rArgThrCysAspArgProAsnCysPro 262
1825 TCAAGTTTCAACAAGCAGGACTGCCCA 1852
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999, DAT: AAX21356
seq_documentation_block:
ID AAX21356 standard; cDNA; 5412 BP.
XX AC
XX AAX21356;
XX AC
XX DT 21-MAY-1999 (first entry)
XX XX
XX Human BAI2 gene.
XX DE
XX Human; BAI2; brain; cancer; drug; diagnosis; prevention; treatment
XX OS Homo sapiens.
XX XX
XX Key Location/Qualifiers
XX CDS 379..5097
XX FT /*tag= a
XX FT /product= "BAI2 protein"
XX PN JP11032766-A.
XX XX
XX XX 09-FEB-1999.
XX XX
XX XX 16-JUN-1997; 97JP-0176485.
XX XX
XX XX 23-MAY-1997; 97JP-0150460.
XX XX (SAKA ) OTSUKA PHARM CO LTD.
XX XX WPI; 1999-183823/16.
XX XX P-PSDB; AAW99301.
XX XX
XX XX New human BAI gene - is expressed in brain plays important role :
XX PT cancer formation
XX PS
XX PS Claim 4; Page 36-42; 62pp; Japanese.
XX CC
XX CC This sequence represents the gene encoding the human BAI2 protein
XX CC new gene is expressed specifically in the brain and plays an impor-
XX CC role in cancer formation in the brain. The BAI2 protein can be u-
XX CC drug compositions to diagnose, prevent or treat such cancers.
XX OS
XX OS Sequence 5412 BP; 943 A; 1790 C; 1652 G; 1027 T; 0 other;

```

```

alignment_scores:      Length: 457
                        Quality: 143.00
                        Ratio: 0.745
                        Gaps: 19
                        Percent Identity: 19.912
alignment_block:
US-09-863-824-2 x AAX21356
..
Align seg 1/1 to: AAX21356 from: 1 to: 5412

```



```

337 llePheaspArgIleLysAargLyAsppHe.....ArgTrpLy 349
    ||| :||| :||| :||| :||| :||| :||| :|||
1927 GCCTTCATGAGATGTGCAGGGATCAGTACGTATGCTGATGACGTGAA 1976

349 saspAlasErGlyProLYsGLuLYsLeuGluiLeYrLYsPrOThrAla. 365
    | ||| :||| :||| :||| :||| :||| :||| :|||
1977 GAAGGCAGCTGCTGGCGAGATCATCTACAACAAGTGCCCCCAAGTCCT 2026

366 .....ArgTYrCYsileArGr..... 371
    ||| :||| :||| :||| :||| :||| :||| :|||
2027 CAGGGTCTCCAGCGGCCGCTGCTCTCTCAGTGCCCAAGCGTGCGGTAC 2076

371 ..... 371

2077 TGGGGCGTCCCAGACTTTGCTGCTGTCATCTCCCATGAGTACCGTACC 2126

372 .MetLeuSerLeuGUSeThrThrLeuAlaAlaglnHisCyScystyrG 388
    ||| ||||||||| :||| :||| :||| :||| :||| :|||
2127 GTATCTGTCACTTAGGGAGCACCTGGCCNAGGGCAGCGCATGCTGGCAG 2176

388 lYaspASnMethlEuilleThrArgGLyLYsGIYAAlaGLyThrProASn 404
    ||| :||| :||| :||| :||| :||| :||| :|||
2177 GCGAGGCGCATGTCGCGAGTGTGGCC.....AGCCTCCAGGAG 2214

405 leuilleSer.....ThrGlUPheSerAlaGluleuHIStyrlysValAs 419
    ||| :||| :||| :||| :||| :||| :||| :|||
2215 CTACTGGCCCGCGCACCTACTATAGTGGGACCTGCTCTTCTGTGGA 2264

419 pVallEUprTpIlleilCYsLYsGIYAstPrSerArgTytRasnClua 436
    | ||| :||| :||| :||| :||| :||| :||| :|||
2265 CATTCGTG.....A 2272

436 laArgProProASNsnGlyGlnLYscYsthrgLUseRProserAspGlu 452
    :||| :||| :||| :||| :||| :||| :||| :|||
2273 GGAATGTCATGCACACCTTTAAGAGGGCCACCTACGTGCCCTCGGCTGAT 2322

453 AspTYrILLELYSglnPhEGln 459
    ||| :||| :||| :||| :||| :||| :||| :|||
2323 GATGTGCAGCGCTTCTCCAG 2343
```

[illegible]

---

OM of: US-09-863-824-2 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Apr 25, 2002 10:16 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framer\_p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09863824/runat\_24042002\_150505\_13882/app\_query.fasta\_1.527  
-DB=Issued\_Patents\_NA -QFWT=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFWT=pfs  
-NORM\_ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09863824@cgn1\_1.89 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-863-824-2  
Query length: 464  
Database: Issued\_Patents\_NA:  
Database sequences: 383533  
Database length: 122816752  
Search time (sec): 50.240000

score\_list:

Sequence	Strd Orig	zscore	Escore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-862-021B-9		130.00	191.20	0.0058	4029
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-313-288B-9		130.00	191.20	0.0058	4029
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-306-902A-3		130.00	191.20	0.0058	4029
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-985-526-4		115.50	178.34	0.0303	1326
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-996-083-2		115.50	164.60	0.1764	4183
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-429-516-2		115.50	164.60	0.1764	4183
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-341-587-4		113.50	157.08	0.4627	5802
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-985-526-2		113.00	182.22	0.0184	657
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-103-840A-2		110.50	72.38	2.4e+04	4403765
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-103-840A-1		110.50	72.36	2.4e+04	4411529
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-808-982-1		107.00	153.19	0.7625	3014
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-306-902A-1		107.00	153.19	0.7625	3014
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-07-862-021B-13		106.50	158.34	0.3936	1816
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-313-288B-13		106.50	158.34	0.3936	1816
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-306-902A-3		106.50	158.34	0.3936	1816
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-07-862-021B-11		105.50	149.67	1.20	3226
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-313-288B-11		105.50	149.67	1.20	3226
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-313-288B-11		105.50	149.67	1.20	3226
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-808-982-3		103.00	146.72	1.75	2831
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-306-902A-3		103.00	146.72	1.75	2831
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-103-840A-1		103.00	58.83	1.3e+05	4411529
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-257-073-14		100.00	147.23	1.64	1725
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-449-645A-14		100.00	140.16	4.05	3116
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-702-367A-14		100.00	140.16	4.05	3116
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-723-534-3		100.00	135.65	7.23	4547
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-459-101A-1		99.50	151.41	0.9577	1128
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-449-285A-6		99.00	152.44	0.8396	960
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-080-255-4		99.00	134.79	8.07	4201
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-465-713-4		99.00	134.79	8.07	4201
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-306-902A-3		99.00	134.79	8.07	4201
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-061-376-4		99.00	122.33	39.89	11907
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-320-559-1		99.00	120.18	52.56	14255
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-327-392-1		99.00	120.18	52.56	14255
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-306-691B-55		99.00	120.18	52.56	14255
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-545-860D-1		99.00	120.18	52.56	14255
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-206-006-1		98.50	145.42	2.06	1601
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-768-301-3		98.50	139.46	4.43	2635
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-428-517-1		98.50	104.05	415.88	50937

/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-447-464-2 - 98.00 129.36 16.20  
/cgn2\_6/ptodata/2/ina/5B\_COMB.seq:US-08-716-679-2 - 98.00 129.36 16.20  
/cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-07-917-722-1 + 97.50 139.30 4.53  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-214-278-6 + 97.50 132.06 11.45  
/cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-068-740A-10 + 97.50 132.06 11.45  
seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-07-862-021B-9  
seq\_documentation\_block:  
; Sequence 9, Application US/07862021B  
; Patent No. 5279966  
; GENERAL INFORMATION:  
; APPLICANT: Jessell, Thomas M  
; APPLICANT: Klar, Avihu  
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/862,021B  
; FILING DATE: 19920405  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 40028  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 664-0525  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4029 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 226..2647  
; US-07-862-021B-9

alignment\_scores:  
Quality: 130.00 Length: 455  
Ratio: 0.722 Gaps: 22  
Percent Similarity: 39.560 Percent Identity: 21.099

alignment\_block:

US-09-863-824-2 x US-07-862-021B-9  
Align seg 1/1 to: US-07-862-021B-9 from: 1 to: 4029  
23 ValLeuArgGlySer.....GlyAlaAlaAspGlyProAspAlaAl 36  
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
1215 ATCTCGACGAGGATCTGTGCACCAAGAGTGTCGCTGGGTCCAGAAAGTGG 1264  
36 aLaGlyAsnAlaSerGlnAlaGlnLeuGlnAsnAsnLeu.AsnValGly 52  
|||||  
1265 TGCAGGA.....CCTAAATTCCTGGGATGCTGGC 1293  
53 SerAspThrThrSerGluThrSerPheSerLeuSerLysGluAlaProAr 69

```
1294  ACAGGACAC.....GGGTGACCTACAGTACCAAA 1325
69  gGluHisLeuAspHisGlnAlaAlaHisGlnProPheProArgProArgp 86
1326  C.....AAGCCACAA 1336
86  heArgGlnGluThrGlyHisProSerLeuGlnArgAspPheProArgSer 102
1337  ITTCCTCAGGAAAAATCCGACCCCTGACTAGTCTGGACCATCCTCAGAGT 1386
103  PheLeuLeuAspLeuProAsnProPheProAsp..... 112
1387  CTTTCTATGAC.....CCGGAAGTGGTCCATCACACAAGT 1424
113  .....LeuSerLysAlaAspIleAsnGlyGln***ProAsnI 125
1425  GGCCAGAGTCGTCATCGAGAGAAATGGCCGGAAGGGAGAGAACAATGCAACA 1474
125  le.....GlnValThrIleGluValValAspGlyProAspSerGlu 138
1475  TTGTACCTGACAATGTGATGATATATTGACCGACCTCGCTCCAGAGAG 1524
139  AlaAspLysAspGlnHisProGlu.....AsnLysProse 150
1525  AAGATGAAGATGACACCCTGAACCTGCATCTACTCCAACTGGTCCCC 1574
150  rTrp..... 151
1575  ATGGTGGCGCTGCAGCTCTTCCACTTGTGAAGGGTAAGAGATGGCGC 1624
152  .....SerValProSerProAsp... 157
1625  AAGCATGTGTGAAGCACAGCTGGACCTCAGTCTCCCTGCTCCTCACACC 1674
157  ..... 157
1675  CAGGACTTCCAGCCCTGCAGTGGCCCGCGCTGCAGCATGAAGATGGCTC 1724
158  .....TrpArgAlaTrpTrpGlnArgSerLeuSerL 168
1725  CACCTGTACCATGTGCGAGTACCTGACCTGACCTGCGTCTGTCTGT 1774
168  euAlaArgAlaAsnSerGlyAspGlnAspTyr***TyrAspSerThrSer 184
1775  GTGGCATGGGTATGAGGTCCCGGAGAGGTACGTGAAGCAGTTCCGGAA 1824
185  AspAspSerAsnProPheLeuAsnProProArgGlyTrpAspHisThrAlaPr 201
1825  GACGGCTCGGTGTGCATGTCTGCC..... 1848
201  oGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerThrA 218
1849  .....ACGGAGACACAGAGAGTGCAGGTCAACGAGGAGTGT 1888
218  spGlyGluGly.....AspTrpSerLeuTrpSerValCysSer 230
1889  CTCTAGCAGTGCCTGCTGACTGAGTGGGGTGGTGGTGGTGGTGGTGGT 1938
231  ValThrCysGlyAsnGlyAsnGlnLysArgThrArg..... 242
1939  GCCACCTGTGGATGGCATGAAGAGAGCCGCCCATCGTCAAGATGAG 1988
243  ....SerCysGlyTyrAlaCysThrAla.....ThrGluSerArgThrC 256
1989  CCCCGCGGAGCGTCCATGTGCAAGGGGAGAGTTCGAGCGGAGAAAT 2038
256  ysAspArgProAsnCysProGlyIleGluAspThrPheArgThrAlaAla 272
2039  GCATGATGCTGTGATGCGCATACCATCCCG..... 2067
273  ThrGluValSerLeuLeuAlaGlySerGluGluPheAsnAlaThrLysLe 289
```

```
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-313-288B-9
seq_documentation_block:
; Sequence 9, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; NAME/KEY: CDS
```

LOCATION: 226..2647  
US-08-313-288B-9

alignment\_scores:

Quality: 130.00 Length: 455  
Ratio: 0.722 Gaps: 22  
Percent Similarity: 39.560 Percent Identity: 21.099

alignment\_block:

US-09-863-824-2 x US-08-313-288B-9 ..

Align seg 1/1 to: US-08-313-288B-9 from: 1 to: 4029

```

23 ValLeuArgGlySer.....GlyAlaAlaAspGlyProAspAlaAl 36
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1215 ATCTGCAGAGGATCTGTCACCAAGGAGTGGCTGGTGGTCAGAAAGTG 1264
36 aAlaGlyAsnAlaSerGlnAlaGlnLeuGlnAsnAsnLeu.AsnValGly 52
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1265 TGCAGGA.....CCTAATTCCCTGGGATGCTGGC 1293
53 SerAspThrThrSerGluThrSerPheSerLeuSerLeuSerGluAlaProAr 69
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1294 ACGGACAGC.....GGGGTACCTACGAGTACCAAA 1325
69 gGluHisLeuAspHisGlnAlaAlaHisGlnProPheProArgProArgP 86
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1326 C.....AAGCCCAAA 1336
86 heArgGlnGluThrGlyHisProSerLeuGlnArgAspPheProArgSer 102
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1337 TTCCTCAGGAATAATCCGACCCCTGACTAGTCTGGACCATCTCCAGAGT 1386
103 PheLeuLeuAspLeuProAsnPheProAsp..... 112
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1387 CCTTCTCTATGAC.....CCGGAAGGTGGTCCATCACACAAGT 1424
113 .....LeuSerLysAlaAspIleAsnGlyCln***ProAsnI 125
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1425 GCCCAGAGTCTCATCGAGAAATGCGCGAAGGAGAAACAATGCAACA 1474
125 le.....GlnValThrIleGluValValAspGlyProAspSerGlu 138
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1475 TTGTACCTGCAGAAATGTGATATTTAGCCGACCTGGCTCCAGAGAG 1524
139 AlaAspLysAspGlnHisProGlu.....AsnLysProSe 150
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1525 AAAGATGAAGATGACACCCCTGAAACCTGCATCTACTCCAACTGGTCCC 1574
150 rTrp..... 151
||| ||||| ||||| ||||| ||||| ||||| |||||
1575 ATGTCGGCCTGCAGCTTCCACTTTGAAAGGTAAGAGGATGGCGC 1624
152 .....SerValProSerProAsp... 157
||| ||||| ||||| ||||| ||||| ||||| |||||
1625 AACGCATGCTGAAGGCACAGCTGGACCTCAGTGTCCCTGCTGCACACC 1674
157 ..... 157
1675 CAGGACTTCCAGCCTGCATGGGCCCGGCTGCAGCGATGAAGATGGCTC 1724
158 .....TrpArgAlaTrpTrpGlnArgSerLeuSerL 168
||| ||||| ||||| ||||| ||||| ||||| |||||
1725 CACCTGTACCATGTCGGAGTGGATCCTGCTCACCCTGCAGTGTCTCGT 1774
168 euAlaArgAlaAsnSerGlyAspGlnAspTyr***TyrAspSerThrSer 184
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1775 GTGGCATGGGTATCAGGTCCCGGAGGAGTACGTGAAGCAGTTCCTCCGAA 1824
185 AspAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAlaPr 201
||| ||||| ||||| ||||| ||||| ||||| |||||
1825 GACGGCTCGGTGTGCTGCTGCC..... 1848

```

```

201 oGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerThrA 218
||| ||||| ||||| ||||| ||||| ||||| |||||
1849 .....ACGGAAGACAGAGAGTGCACGGTCAACGAGGAGTCT 1888
218 spGlyGluGly.....AspTrpSerLeuTrpSerValCysSer 230
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1889 CTCCTAGCAGCTGCCTGGTGTGACTGAGTGGGTGAGTGGGATGACTGCAGC 1938
231 ValThrCysGlyAsnGlyAsnGlnLysArgThrArg..... 242
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1939 GCCACCTGTGAATGGGCATGAAGAAGCGGCACCGCATGGTCAAGATGAG 1988
243 ....SerCysGlyTyrAlaCysThrAla.....ThrGluSerArgThrC 256
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1989 CCCCGGCGACGCTCCATGTGCAAGGCGAGACTTCGCGAGCGGAGAAAT 2038
256 ysAspArgProAsnCysProGlyIleGluAspThrPheArgThrAlaAla 272
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2039 GCATGATGCCTGAGTGCATACCATCCCG..... 2067
273 ThrGluValSerLeuAlaGlySerGluGluPheAsnAlaThrLysLe 289
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2068 .....TGCTTGTCTCTCTGTGTCGAGTGGAGCGACTGTAGCGT 2108
289 uPheGluValAspThrAspSerCysGluArgTrpMetSerCysLysSerG 306
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
2109 G.....ACCTGTGGGAAGGCGATCGGAGCGCCAGC 2140
306 luPheLeuLys.....Lys 310
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
2141 GGATGCTCAAGTCTCTGCGCAGAGTGGGGACTGTAATGAGGATCTGGAG 2190
311 TyrMetHisLysValMetAsnAspLeuProSerCysPro.....CysSe 325
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
2191 CAGCGGAGAAAGTGTATG.....CTGCCAGAGTCCCGCATTTACTGCGGA 2234
325 rTyProThrGluValAlaTyrSerThrAlaAsp..... 336
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2235 ACTCAGTGAGTGGTCCCGAGTGTCTGAATGTAAACAAGTCTGTGGGAAG 2284
337 .....IlePheAspArgIleLysArgLysAspPheArgTrpLysAsp 350
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2285 GTCACATGATCGAACCAGGCAATCCAAATGGAACCTCAGTTGGAGGT 2334
351 AlaSerGlyProLysGluLysLeuLysLeuLysProThrAlaArgTy 367
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2335 GCACCCCTGCCCA...GAGACTGTGCACGCAAGAAGTCCCGTCCCGGAA 2381
367 rCysIleArgSer 371
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2382 ATGCTTCGCAGC 2394

```

seq\_name: /cgn\_6/ptodata/2/lna/PCTUS\_COMB.seq:PCT-US93-03164-9

seq\_documentation\_block:

```

; Sequence 9, Application PC/TUS9303164
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/03164
: FILING DATE: 19930402
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 40028
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 977-9550
: TELEFAX: (212) 664-0525
: TELEX: 422523 COOP UI
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4029 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 226..2647
: PCT-US93-03164-9

alignment_scores:
  Quality: 130.00 Length: 455
  Ratio: 0.722 Gaps: 22
Percent Similarity: 39.560 Percent Identity: 21.099

alignment_block:
US-09-863-824-2 x PCT-US93-03164-9 ..
Align seg 1/1 to: PCT-US93-03164-9 from: 1 to: 4029

23 ValLeuArgGlySer.....GlyAlaAlaAspGlyProAspAlaAl 36
::: |||||
1215 ATCTCGAGAGGATCTGTCGACCAAGAGGTGGCTGGTCCAGAAAGTGG 1264
::: |||||
36 aAlaGlyAsnAlaSerGlnAlaGlnLeuGlnAsnLeu..AsnValGly 52
|||
1265 TGCAGGA.....CCTAATCCCTGGGATGCTGCG 1293
|||
53 SerAspThrThrSerCluThrSerPheSerLeuSerLysGluAlaProAr 69
::: |||||
1294 ACGGACAGC.....GGGGTGACCTACGATCACCACAA 1325
|||
69 gGluHisLeuAspHisGlnAlaAlaHisGlnProPheProArgProArgP 86
::: |||||
1326 C.....ANGCCACAA 1336
|||
86 heArgGlnGluThrGlyHisProSerLeuGlnArgAspPheProArgSer 102
::: |||||
1337 TTCTCAGGAAAAATCCGACCCCTGACTAGTCTGGACCATCTCAGAGT 1386
|||
103 PheLeuLeuAspLeuProAsnPheProAsp..... 112
::: |||||
1387 CCTTCTATGAC.....CCGGAAGGTGGTCCATCACACAAGT 1424
|||
113 .....LeuSerLysAlaAspIleAsnGlyGln***ProAsnI 125
::: |||||
1425 GGCCAGAGTCTGTCGAGAGAAATGCCCGAAGGAGGAGAACAAATGCAACA 1474
|||
125 le.....GlnValThrIleGluValValAspGlyProAspSerGlu 138
|||
1475 TTGTACCTGACAATGTGGATGATATTGTAGCCGACCTGCTCCAGAAAGAG 1524
|||
139 AlaAspLysAspGlnHisProGlu.....AsnLysProSe 150

```

```

|||:::||||::: |||||
1525 AAAGATGAAGATGACACACCCCTGAAACCTGCATCTACTCCAACGTGGTCCCC 1574
|||
150 rTrp..... 151
|||
1575 ATGGTGGCCTGCAGCTCTTCCACTTGTGAAAAGGGTAAAGAGGATCGGC 1624
|||
152 .....SerValProSerProAsp... 157
|||
1625 AACGCATGCTGAAGGACACAGCTGAGCTCAGTGTCCCTCTCCTGCACACC 1674
|||
157 ..... 157
1675 CAGGACTTCCAGCCCTGCATGGGCCCGGCTGCAGGATGAAGATGGCTC 1724
158 .....TrpArgAlaTrpTrpGlnArgSerLeuSerL 168
|||
1725 CACCTGTACCATGTCGGAGTGGATCACCTGGTCACCTGCAGTGTCTCGT 1774
|||
168 euAlaArgAlaAsnSerGlyAspGlnAspTyr**TyrAspSerThrSer 184
::: ::: ::: |||
1775 GTGGCATGGGTATGAGGTCCCGGAGAGGTACGTGAAGCAGATTCCCGGAA 1824
|||
185 AspAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAlaPr 201
|||
1825 GACGGCTCGGTGTCATGCTGCC..... 1848
|||
201 oGlyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerThrA 218
|||
1849 .....ACGGAAGACAGACAGAGAGTGCACGGTCAACGAGGAGTGTCT 1888
|||
218 spGlyGluGly.....AspTrpSerLeuTrpSerValCysSer 230
::: ::: |||||
1889 CTCCTAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1938
|||
231 ValThrCysGlyAsnGlyAsnGlnLysArgThrArg..... 242
::: |||||
1939 GCCACCTGTGAATGGGATCAAGAGCGGCACCGCATGGTCAAGATGAG 1988
|||
243 ....SerCysGlyTyrAlaCysThrAla.....ThrGluSerArgThrC 256
::: ||| ||| |||
1989 CCCCGCGCAGCGCTCCATGTGTCAAGAGCGGAGACTTCGCGAGCGGAGAAAT 2038
|||
256 ysAspArgProAsnCysProGlyIleGluAspThrPheArgThrAlaAla 272
|||
2039 GCATGATGCCCTGATGCCATACCATCCG..... 2067
|||
273 ThrGluValSerLeuLeuAlaGlySerGluGluPheAsnAlaThrLysLe 289
|||
2068 .....TGCTTGCTGCTCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 2108
|||
289 uPheGluValAspThrAspSerCysGluArgTrpMetSerCysLysSerG 306
::: |||
2109 G.....ACCTGTGGGAAGGGCATCGCGACGCGCCGAGC 2140
|||
306 luPheLeuLys.....Lys 310
::: |||||
2141 GGATGCTCAAGTCTCTGCGACAGCTGGGGGACTGTATGAGATCTGGAG 2190
|||
311 TyrMetHisLysValMetAsnAspLeuProSerCysPro.....CysSe 325
|||
2191 CAGGCGGAGAAAGTGTATG.....CTGCCAGAGTGCCTGCTGCTGCTGCTG 2234
|||
325 rTyrProThrGluValAlaTyrSerThrAlaAsp..... 336
::: |||||
2235 ACTAGTGAAGTGGTCCCGAGTGTCTGAATGTAACCAAGTCTCTGTGGGAAAG 2284
|||
337 .....IlePheAspArgIleLysArgLysAspPheArgTrpLysAsp 350
|||
2285 GTCACATGATTGGAACCCCGACAATCCAAAATGGAACCTCAGTTTGGAGGT 2334
|||
351 AlaSerGlyProLysGluLysLeuGluLysProThrAlaArgTy 367
|||

```



seq name: /cgn2 6/otodata/2/ina/6B COMB.seq:US-09-429-516-2



```

190 euAsnProProArgGly...Tip.....AspHisThrAlaProGly... 202
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
395 TGAACCCACCGCGGCTTCGGTGGCTCAACCGGAGCAACCGCGTGGC 444
203 .....HisArgThrPheGluThrLysAspGlnProGln 213
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
445 CGCGCGCTCCTCAACTACCACTGCGCTTCGGCTGCCCACTAGAACGCTC 494
213 uTyAspSerThrAspGlyGluGlyAspTrpSerLeuTrpSerValCysS 230
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
495 GTGG.....GGCGGTGGGCGCGGTGGGCTCCCTGCT 526
230 erValThrCysGlyAsnGlnLysArgThrArgSerCys..... 244
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
527 CGGGAGCTGTGGCCAGCGCTGCTGCGCGCGCGCCCACTGCCCAAGC 576
245 .....GlyTyrAlaCysThrAla.....ThrGluSerArgThrCysAs 257
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
577 CCCGCTGGGATGGCTGCTCCGCGCGCTCCTCTGGAGGCGCAGAAAGTGGT 626
257 pArgProAsnCysProGly 263
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
627 CGGCGCTGGGTGTCAGGG 645

```

seq\_name: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:US-09-341-587-4

```

seq_documentation_block:
; Sequence 4, Application US/09341587
; Patent No. 6346606
; GENERAL INFORMATION:
; APPLICANT: Mollenhauer, Jan
; TITLE OF INVENTION: Protein Containing an SRCR Domain
; FILE REFERENCE: 4121-108
; CURRENT APPLICATION NUMBER: US/09/341,587
; CURRENT FILING DATE: 1999-08-31
; EARLIER APPLICATION NUMBER: PCT/DE98/00096
; EARLIER FILING DATE: 1998-01-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 5802
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-341-587-4

```

```

alignment_scores:
Quality: 113.50      Length: 230
Ratio: 1.102        Gaps: 15
Percent Similarity: 44.783      Percent Identity: 26.522

alignment_block:
US-09-863-824-2 x US-09-341-587-4  ..
Align seg 1/1 to: US-09-341-587-4 from: 1 to: 5802

143 GlnHisProGluAsn.....LysProSerTrpSe 152
||||| ||||| ||||| ||||| ||||| ||||| |||||
1463 CAGCACAGTGAAGACGCTGGTGTCTCTCTGCTGCCCACTCTCTGCTC 1512
152 rValProSerProAspTrp.....ArgA 160
|:||||| ||||| ||||| ||||| ||||| ||||| |||||
1513 GACGCCAGTCCAGACACATGCGCACCATCACCTTGGCTGCATCGACAG 1562
160 laTrpTrpGlnArgSerLeuLeuAlaArgAlaAsnSerGlyAspGln 176
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1563 TAGGATCTGAATCCAGTTTGGCCCTGAGGCTGGTGAATGGAGTGACAGG 1612
176 ..... 176
1613 TGTCAGGCGCGAGTGGAGGTCTCTATACCAAGGCTCTTGGGCGCAGGTGTG 1662

```

```

177 .AspTyr**TyrAspSerThrSerAspSerAsnPheLeu.....A 191
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1663 CGATCAGAGCTGGGAC.....ACCAATGATGCCAATGCTCTGCAGGC 1706
191 snProProArgGlyTrpAspHisThrAlaProGlyHisArgThrPheGlu 207
::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1707 AACCGGGCTGTGGTGGGCCATGTCAGGCCAGGAAATGCCCGTTGGT 1756
208 ThrLysAspGlnPro...GluTyAspSerThrAspGlyGluGlyAspTr 223
::: ||| ||||| ||||| ||||| ||||| ||||| |||||
1757 CAGGCTCAGGACCACTTCTCTGGATGATGTCGCTCTCAGGACACGA 1806
223 pSerLeu...TrpSerVal.....CysSerValThrCysG 234
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1807 GTCTTACCGTGGAGCTGCCCCACCAATGGCTGCTCTCCACCAACTGTG 1856
234 LysnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAla 250
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1857 GCCAT.....AGTGAAGACGCTGTGTCTCATCTGCTCAGCT 1891
251 ThrGluSerArgThrCysAspArgProAsnCysProGlyIleGluAspTh 267
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1892 TCCAGTCCCGGCCAACACCTAGTCCA.....GACAC 1923
267 rPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerGluGluP 284
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1924 TTGCCAACCTCACATGCATCAACA.....GCAGATCTGAATCCA 1964
284 heAsnAlaThrLysLeuPheGluValAspThrAspSerCysGlu...Arg 299
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1965 GTTTGGCCCTGAGCTGGTGAAT...GGAGGTGACAGGTGTGAGGCGCGA 2011
300 .TrpMetSerCysLysSerGluPheLeuLysLysTyrMethHisLysValM 316
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2012 GTGGAGGTCC.....T 2022
316 eAsnAspLeuProSerCysProCysSerTyrProThr 328
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
2023 ATACCGAGGCTCTGGGCGCACCGTGTGTGATGACTACT 2060
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-985-526-2

```

```

seq_documentation_block:
; Sequence 2, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorrow Jr., Robert G
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 657 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-985-526-2

alignment_scores:
  Quality: 113.00      Length: 74
  Ratio: 2.825        Gaps: 3
  Percent Similarity: 54.054  Percent Identity: 33.784

alignment_block:
US-09-863-824-2 x US-08-985-526-2 ..
Align seg 1/1 to: US-08-985-526-2 from: 1 to: 657
193 ProArgGlyTrpAspHisThrAlaProGlyHisArgThrPheGluThrLy 209
||||| |||
208 CCTCGTGTGG..... 219
209 saspGlnProGluTyAspSerThrAspGlyGluGlyAspTrpSerLeut 226
||||| : : : : : ||||| : : : : : |||||
220 .....CCACGGACTCGCGGAGATGGTGTCTCCATGGTCCGAGT 262
226 rpSerValCysSerValThrCysGlyAsnGlyAsnGlnLysArgThrArg 242
||||| : : : : : ||||| : : : : : |||||
263 GGACCTCCTGTTCTACGAGCTTGGCAATGGAATTCAGCAGCGCGCGGC 312
243 SerCys.....GlyTyralaCys.....ThrAlaThrGluSerAr 254
||||| : : : : : ||||| : : : : : |||||
313 TCCTGGATAGCTCAACACCGATGTGAGGCTCCTCGTCCAGACAG 362
254 gThrCysAspArgProAsnCys 261
||||| : : : : : |||||
363 GACCTGCCACATTCAGGAGTGT 384

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-103-840A-2

seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-08-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
;
US-09-103-840A-2

alignment_scores:
  Quality: 110.50      Length: 421
  Ratio: 0.639        Gaps: 23
  Percent Similarity: 41.093  Percent Identity: 22.090
```

```

alignment_block:
US-09-863-824-2 x US-09-103-840A-2 ..
Align seg 1/1 to: US-09-103-840A-2 from: 1 to: 4403765
31 AspGlyProAspAlaAlaGly.....AsnAlaSerGlnAlaGl 44
: : : : : ||||| : : : : : |||||
30705 GAAGGCGCCGACGAGGACGCGGGCTGGCAAGTCAGAATCGCCACCGCA 30754
44 nLeuGlnAsnAsnLeuAsnValGlySerAspThrThrSerGluThrSerP 61
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
30755 GGCTCCGACGACCGTGACGCTG.....CCCAACGGTGACGCGTGA 30795
61 heSerLeuSerLysGluAlaProArgGluHisLeuAspHisGlnAlaAl 77
: : : : : ||||| : : : : : ||||| : : : : : |||||
30796 CCGGGCCAGTCCCGACGCTCGCGGCGCATCAAGGCGGCGGCACGCGC 30845
77 aHisGlnProPheProArgPheArgPheArgGlnGluThrGlyHisProS 94
||||| : : : : : ||||| : : : : : |||||
30846 ACACCGATCGCAGATGCTTCCACACACAGGAA.....TTGCCAT 30886
94 erLeuGlnArgAspPheProArgSerPheLeuLeuAspLeuProAsnPhe 110
||||| : : : : : ||||| : : : : : |||||
30887 CCGCTACCGGAA.....CCGCGGTGC.....CCACCCGTCGACT 30924
111 ProAspLeuSerLysAlaAspIleAsnGlyGln***ProAsnIleGlnVa 127
||||| : : : : : ||||| : : : : : |||||
30925 CCG..... 30927
127 lThrIleGluValValAspGlyProAspSerGluAlaAspLysAspGlnH 144
||||| : : : : : ||||| : : : : : |||||
30928 .....CCGGATCTCAGCGGGAGACGAGTAGTGT 30955
144 isProGluAsnLysPro...SerTrp.....SerValProSer 155
: : : : : ||||| : : : : : |||||
30956 GTTCACCGCCAGCGCTTGCCCTTGCCCTAGCAAGCTCTTCTGGACGG 31005
156 ProAspTrpArgAlaTrpTrpGlnArgSerLeuSerLeuAlaArgAla.. 171
||||| : : : : : ||||| : : : : : |||||
31006 CCAGAT.....TCAACACATCTCAGCCGTGGGAGGCCAA 31040
172 .....AsnSerGlyAspGlnAspTyr***TyrAsps 182
: : : : : ||||| : : : : : ||||| : : : : : |||||
31041 ACTTCTAGGCTGGATATATCCAGCGCGACCGCGCCGCGCGGAGG 31090
182 erThrSerAspAspSerAsnPheLeuAsnProProArgGlyTrpAspHis 198
: : : : : ||||| : : : : : |||||
31091 ACCGAAGCACCGACACCAAC..... 31110
199 ThrAlaProGlyHisArgThrPheGluThrLysAspGlnProGluTyraS 215
||||| : : : : : ||||| : : : : : |||||
31111 ...CAGGCGCGCGCTCGATAGGTACT...GACCGCCCGGTC..... 31149
215 pSerThrAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerVal. 231
: : : : : ||||| : : : : : ||||| : : : : : |||||
31150 .ACAACAGAGAGACAGCGGATGACAGATCGAATTCAGTCGACGCTGC 31198
232 .....ThrCys 233
||||| : : : : : |||||
31199 ACATTTAGTTCAGGCGCTGCCCATCACCAGCAGACCGCGCTACCTGC 31248
234 GlyAsnGlyAsnGlnLysArgThrArgSerCys..... 244
: : : : : ||||| : : : : : ||||| : : : : : |||||
31249 GGACCGTCCGTCGTCGCGACGACGCGATCCGGAAAGTCTGGACTCGCTG 31298
245 .....GlyTyralaCysThrAlaT 251
||||| : : : : : ||||| : : : : : |||||
31299 GGGCCTATTTTCAGTCAGTCCGCGACACCGCGCTGAGTCTCGAGCT 31348
251 hrGluSer..... 253
: : : : : |||||
31349 CAGAAAGCAGTCTCTACCAGCAAGCGCGCAACACCGCGGTATTTGCC 31398
```

```

alignment_scores:
  Quality: 110.50      Length: 421
  Ratio: 0.639        Gaps: 23
  Percent Similarity: 41.093  Percent Identity: 22.090

alignment_block:
  US-09-863-824-2 x US-09-103-840A-1 ..

  Align seg 1/1 to: US-09-103-840A-1 from: 1 to: 4411529

31 AspGlvPvAspAlaAlaGlv.....AsnAlaSerGlnAlaGln 44

```







```

: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/313,288B
: FILING DATE: January 5, 1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0526
: TELEX:
:
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1816 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2..1705
: US-08-313-288B-13
:
: alignment_scores:
: Quality: 106.50 Length: 329
: Ratio: 0.813 Gaps: 14
: Percent Similarity: 39.818 Percent Identity: 21.277
:
: alignment_block:
: US-09-863-824-2 x US-08-313-288B-13 ..
:
: Align seg 1/1 to: US-08-313-288B-13 from: 1 to: 1816
:
: 93 ProSerLeuGlnArgAspPheProArgSerPheLeu...LeuAspLeuPr 108
: |||.....:||||| |||...:||||| ||
: 386 CCCACCATCCCGAGGATTAATCCGACCTCTGACAGTCTGGGATCACCC 435
:
: 108 oAsnPheProAspLeuSerLysAlaAspIleAsnGlyGln***ProAsnI 125
: ||: |||:||||:||||: ||: ||
: 436 ACAAAGCCCTTCTATGACAGAGT.....GGCCCAATCATACCTATAG 479
:
: 125 leGlnValThrIleGlu..... 130
: :|||:|||||
: 480 CTCGAGTTGTGATTGAAGGATTGCCAGGAGGAGAACAGTCAATATT 529
:
: 131 .....ValValAspGlyProaspSerGluAl 139
: |||:|||||
: 530 ATACCCGACAACTGGATGACATAGTACAGATCTGTGTAACGGAAGAGAA 579
:
: 139 aAspLysAspGlnHisProGlu.....AsnLysProSerT 151
: |||:||||: |||: |||
: 580 AGACAAGATGATACCCCGGAGACCTGCATATATTCCAACTGGTCCCT 629
:
: 151 rp..... 151
: ||
: 630 GGTGGCTGTCAGCTGCGCCACCTGCGACAAAGGCGGATGAGACAG 679
:
: 152 .....SerValProSerProasp..... 157
: |||:|||||
: 680 CGCATGTTAAAGGCTCAGTTAGATCTCAGTGTTCCTGCCCGACACTCA 729
:
: 157 ..... 157

```

```

730 GGACTTTTAAACCTGCATGGGCGCGGTGCAGCGATGACGAGCCCTCTA 779
158 .....TrpArgAlaTrpTrpGlnArgSerLeuSerLeu 168
: ||| :||| ||| |||
780 CCTGCATGATGTCAGAAATGGATCACCTGGTGCAGCGCTCCTCTGC 829
169 AlaArgAlaAsnSerGlyAspGlnAspTyr***TyrAspSerThrSerAs 185
: ||| :||| ||| |||
830 GGGATGGGAATTGAGTGCAGGAGAGATACGTCAAGCAGTTCCAGAGAA 879
185 pAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAlaProG 202
: ||| |||
880 CGGTTCTCTGTGTAAGTCCCA..... 901
202 lyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerThrAsp 218
: ||| |||: ||| :|||
902 .....ACGGAAGAACTGAGAAATGCATTTGTCAATGAGGAGTGTGAG 943
219 GlyGluGly.....AspTrpSerLeuTrpSerValCysSerVa 231
: ||| :||| ||| |||
944 CCAAGCAGCTGTATAGTCACGGAATGGCAGAGTGGGAGGAGTGCAGCGC 993
231 lThrCysGlyAsnGlyAsnGlnLysArg.....ThrA 242
: ||| ||| :||| |||
994 TACATGCGGATGGTATGAAGAAGCGGCACAGGATGATAAAGATGACTC 1043
242 rgSerCysGlyTyrAlaCysThrAla.....ThrGluSerArgThrCys 256
: ||| ||| ||| |||
1044 CAGCGGATGGATCTATGTGCAAAAGCCGACACACAGAGGTTGAGAAATGC 1093
257 AspArgProAsnCysProGlyIleGluAspThrPheArg.....ThrAl 271
: ||| ||| :||| |||
1094 ATGATGCCGAATGTATACCATCCGTGGTGTGTGCCCTTGTCTCTGA 1143
271 aAlaThrGluValSerLeuLeuAlaGlySerGluGluPheAsnAlaThrL 288
: ||| ||| :||| |||
1144 ATGGAGTGATTGTCAGCGTTACCTGTGCAAAAGCCACCAAGACACAGACA 1193
288 ysLeuPheGluValAspThrAspSerCysGluArgTrpMetSerCysLys 304
: ||| ||| :||| |||
1194 GAATGTTGAAGTCCCGTCTGAACTTGGAGAT.....TGCAAT 1231
305 SerGluPheLeuLysLysTyrMetHisLysValMetAsnAspLeuProse 321
: ||| ||| :||| |||
1232 GAGAACTGGAACTGAAACAAGTGGAAAGTGCATG.....CTTCTCTGA 1275
321 rCysProCysSerTyrPro...ThrGluValAlaTyr 332
: ||| ||| :||| |||
1276 ATGCCCTATAAGCTGTGAATTGACAGAGTGTGTCTTAC 1312
:
: seq_name: /cqn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-03164-13
:
: seq_documentation_block:
: Sequence 13, Application PC/TUS9303164
: GENERAL INFORMATION:
: APPLICANT: Jessell, Thomas M
: APPLICANT: Klar, Avihu
: TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
: TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham
: STREET: 30 Rockefeller Plaza
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/03164
FILING DATE: 19930402
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 40028
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1816 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1705
PCT-US93-03164-13

```

```

alignment_scores:
Quality: 106.50 Length: 329
Ratio: 0.813 Gaps: 14
Percent Similarity: 39.818 Percent Identity: 21.277

```

alignment\_block:

```

US-09-863-824-2 x PCT-US93-03164-13 ..
Align seg 1/1 to: PCT-US93-03164-13 from: 1 to: 1816

93 ProSerLeuGlnArgAspPheProArgSerPheLeu...LeuAspLeuPr 108
|||||:|||||:|||||:|||||:|||||:|||||:
386 CCACCATTCCTCCAGGATAAATCCGCTCTGACAAGTCTGGATCACC 435
108 oAsnPheProAspLeuSerLysAlaAspIleAsnGlyGln***ProAsnI 125
|||||:|||||:|||||:|||||:|||||:
436 ACAAGCCCTCTATGACCAGAGGT.....GGGCCAATCATACCTATAG 479
125 leGlnValThrIleGlu..... 130
480 CTCAGTTGTGATTGAAGGATTGCCAGGAGGACACAGTGCATATT 529
131 .....ValValAspGlyProAspSerGluAl 139
530 ATACCCGACACCTGGATGACATAGTAGCAGATCTGTTAACGGAAGAA 579
139 aspLysAspGlnHisProGlu.....AsnLysProSerT 151
|||||:|||||:|||||:|||||:
580 AGACGAAGATGATACCCCGAGACCTGCATATATTCCAACTGTGTCCTCC 629
151 rp..... 151
630 GGTCGGCCCTGCAGCTCGGCCACCTGGCAGCAAGGGCAGCGGATGACAG 679
152 .....SerValProSerProAsp..... 157
680 CGCATGTTAAAGGCTCAGTTAGATCTCAGTGTTCCTCCCTGCCAGACACT 729
157 ..... 157
730 GGACTTTGAACCTGTCATGGGCGCGCTGCACGCGATGACGAAGCCTCTA 779
158 .....TrpArgAlaTrpTrpGlnArgSerLeuSerLeu 168
|||||:|||||:|||||:|||||:
780 CCTGCATGATGTCAGATGATGATACCTGCTCGCGCTGCAGCGCTCCTGC 829

```

```

169 AlaArgAlaAsnSerGlyAspGlnAspTyr***TyrAspSerThrSerAs 185
|||:|||:|||:|||:|||:|||:
830 GGGATGGGAATTGAGGTCAGGGAGAGATACGTCACGACGATGTCACAGA 879
185 pAspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAlaProG 202
|||:|||:|||:|||:|||:
880 CGGTTCTCTTGTGTAAGTCCCA..... 901
202 lyHisArgThrPheGluThrLysAspGlnProGluTyrAspSerThrAsp 218
|||:|||||:|||||:|||||:
902 .....ACGGAAGAAACTGAGAAATGCATTCTCAATGAGGAGTGTGAG 943
219 GlyGluGly.....AspTrpSerLeuTrpSerValCysSerVa 231
|||:|||||:|||||:|||||:
944 CCAAGCAGCTGTATAGTCACGGAATGGCAGAGTGGGAGGAGTGCAGCG 993
231 lThrCysGlyAsnGlyAsnGlnLysArg.....ThrA 242
|||||:|||||:|||||:|||||:
994 TACATCCGGATGGGTATGAAGAGCGGCACAGGATGATAAAGATGACTC 1043
242 rgSerCysGlyTyrAlaCysThrAla.....ThrGluSerArgThrCys 256
|||||:|||||:|||||:|||||:
1044 CAGCGATGGATCTATGTGCAAGCCGACACACAGAGTTGAGAAATGC 1093
257 AspArgProAsnCysProGlyIleGluAspThrPheArg.....ThrAl 271
|||||:|||||:|||||:|||||:
1094 ATGATGCCCGAATGTCATACCATCCCGTGGTGTGTCCTCCCTTGGTCTGA 1143
271 aAlaThrGluValSerLeuLeuAlaGlySerGluGluPheAsnAlaThrL 288
|||||:|||||:|||||:|||||:
1144 ATGGAGTGATTTGCAGCGTTACCTGTGCAAGAGGCCACAGAACCCAGACA 1193
288 ysLeuPheGluValAspThrAspSerCysGluArgTrpMetSerCysLys 304
|||||:|||||:|||||:|||||:
1194 GAATGTTGAAGTCCCGCTGCAACTTGGAGAT.....TGCAAT 1231
305 SerGluPheLeuLysLysTyrMetHisLysValMetAsnAspLeuProSe 321
|||||:|||||:|||||:|||||:
1232 CAGGAACCTGGAACTGAAACACAGAGTGGAAAGTGCATG.....CTTCCTGA 1275
321 rCysProCysSerTyrPro...ThrGluValAlaLys 332
|||||:|||||:|||||:|||||:
1276 ATGCCCTATAAGCTGTGNATTGACAGAGTGGTCTTAC 1312

```



OM of: US-09-863-824-2 to: EST:\* out\_format : pfs  
Date: Apr 25, 2002 9:43 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODE=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USP70\_spool/US09863824/runat\_24042002\_150505\_13849/app\_query.fasta\_1.527  
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.500  
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000  
-GAPEXT=7.000 -XGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pls -NORM=ext -HEAPSIZ=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09863824@cgn1\_1.4938  
-NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
-NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-863-824-2

Query length: 464

Database: EST:\*

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 1532.850000

score\_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_est2:BF472502	-	1103.00	1790.72	1.7e-90	644	1 BE472502 UI-M-BH3-axd-h-03-0-UI
gb_est2:BG162342	+	1077.00	1747.41	4.4e-88	691	1 BG162342 de3se08.yl Wellcome CR
gb_est1:AW873508	+	1034.00	1711.67	4.3e-86	583	1 AW873508 ho0h09.xl Soares_NFL
gb_est2:BJ060213	+	968.00	1569.90	3.4e-78	645	1 BJ060213 BJ060213 NIBB Mochil
gb_est2:BF302850	+	854.00	1378.97	1.5e-67	968	1 BF302850 602032745F1 NCI_CGAP
gb_est2:BE947736	+	827.50	1341.99	1.7e-65	544	1 BE947736 UI-M-BH3-axd-h-03-0-UI
gb_est2:BT891452	+	703.50	1136.96	4.4e-54	666	1 BT891452 2F637-3-000707 Zebra
gb_est1:AA884248	+	684.50	1110.84	1.3e-52	425	1 AA884248 ac32b09.sl Soares_NFL
gb_est2:BT170419	+	678.50	1095.28	9.2e-52	715	1 BT170419 zfsTARAMA000029 Zebra
gb_est1:AW426875	+	671.00	1089.91	1.8e-51	383	1 AW426875 61687 MARC 3BOV Bos ta
gb_est2:BM191035	+	619.00	1003.18	1.2e-46	446	1 BM191035 daj31b08.yl NICHD XGC
gb_est2:WB6257	+	594.00	964.78	1.7e-44	356	1 WB6257 zdg91f09.r1 Soares fetal
gb_est1:AW455535	+	574.50	932.20	1.1e-42	379	1 AW455535 UI-M-BH3-avb-e-02-0-UI
gb_est2:BT533390	+	564.00	910.93	1.7e-41	549	1 BT533390 398882 MARC 4BOV Bos U
gb_est1:AL545675	+	525.00	840.18	1.5e-37	1032	1 AL545675 AL545675 LTI_NFL006.F
gb_est1:AW435642	+	496.00	801.23	2.2e-35	479	1 AW435642 74577 MARC IPTG Sus sc
gb_est2:BF571346	+	491.00	789.15	1.0e-34	682	1 BF571346 602077485F1 NIH_MGC.62
gb_gss:CN5021K6	+	485.00	776.92	5.0e-34	914	1 AL199023 Tetraodon nigroviridis
gb_est1:BG037854	+	478.50	772.98	8.3e-34	463	1 BG037854 dc69a05.yl NICHD XGC H
gb_est1:BB322233	+	472.00	758.72	5.2e-33	643	1 BB322233 BB322233 RIKEN full-le
gb_est1:AL659826	+	467.00	749.94	1.6e-32	679	1 AL659826 AL659826 XGC-neurula S
gb_est1:AW435639	+	449.00	730.85	1.8e-31	266	1 AW435639 74571 MARC IPTG Sus sc
gb_est2:W14682	+	354.50	575.93	7.8e-23	275	1 W14682 mb30h04.r1 Soares mouse
gb_est2:BJ071721	+	342.00	546.72	3.3e-21	608	1 BJ071721 BJ071721 NIBB Mochil
gb_est2:BG739780	+	337.00	535.15	1.5e-20	827	1 BG739780 602630502F1 NCI_CGAP
gb_est1:AA447294	+	327.50	520.74	9.3e-20	747	1 AA447294 AC.E04R PH001-Normal H
gb_est2:BE287313	+	308.00	489.67	5.0e-18	693	1 BE287313 601093984F1 NCI_CGAP.M
gb_est2:BT1442837	+	277.00	441.60	2.4e-15	546	1 BT1442837 daj91a08.yl NICHD XGC
gb_est1:AW435643	+	256.00	422.20	2.9e-14	141	1 AW435643 74579 MARC IPTG Sus sc
gb_gss:AO606601	+	255.50	404.13	2.9e-13	673	1 AO606601 HS.2095.B1.E07 MR CIT
gb_est1:BB586477	+	242.50	391.39	1.5e-12	311	1 BB586477 BB586477 RIKEN full-le
gb_est1:AW435648	+	213.00	352.52	2.2e-10	133	1 AW435648 74587 MARC IPTG Sus sc
gb_est2:BU082998	-	211.00	330.88	3.5e-09	702	1 BU082998 BJ082998 NIBB Mochil
gb_est2:BT1449830	+	206.00	328.04	5.0e-09	433	1 BT1449830 dag32a03.y3 Blackshea
gb_gss:AZ995663	+	168.50	263.31	2.0e-05	589	1 AZ995663 2M0281N06R Mouse 10kb
gb_est1:AT1430783	-	150.00	240.58	0.0004	298	1 AT1430783 hb66a04.xl Soares mous
gb_est1:AT1070670	-	140.00	222.35	0.0039	353	1 AT1070670 UI-R-C2-mx-c-05-0-UI.S
gb_est1:AA464532	+	131.50	203.45	0.0437	555	1 AA464532 zx85a05.sl Soares ova
gb_est2:H24875	-	129.00	212.30	0.0141	172	1 H24875 y142h10.r1 Soares breast
gb_est1:AA243836	-	129.00	198.78	0.0796	585	1 AA243836 zr67g12.r1 Soares_NHNM

gb\_gss:AG041438 - 128.50 196.94 0.1008 642 1 AG041438 Pan troglodytes DN  
gb\_est2:BF769031 - 128.00 199.20 0.0755 486 1 BF769031 ILO-IT0033-101100-1  
gb\_est1:AA243572 + 127.50 195.91 0.1151 608 1 AA243572 zr67g12.sl Soares  
gb\_est1:AT278172 + 127.00 198.66 0.0809 440 1 AT278172 ql77h07.xl Soares  
gb\_est1:AI866968 + 127.00 197.95 0.0885 469 1 AI866968 wnl3f10.xl NCI\_CGA

seq\_name: gb\_est2:BF472502

seq documentation block:

LOCUS BF472502 644 bp mRNA linear EST 04-DEC-2000

DEFINITION UI-M-BH3-axd-h-03-0-UI.r1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone

UI-M-BH3-axd-h-03-0-UI 5', mRNA sequence.

ACCESSION BF472502

VERSION BF472502.1 GI:11541685

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 644)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL 97044477

MEDLINE

COMMENT

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mEst@mail.nih.gov

cDNA Library Preparation: M.B. Soares Lab Clone Distribution:

Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It

should be noted that Bento Soares is generating a small number of

additional specialized non-redundant arrays of BMAP cDNAs whose

availability will be considered under appropriate and limited

collaborative arrangements

Seq primer: M13 Reverse.

Location/Qualifiers

1..644

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UI-M-BH3-axd-h-03-0-UI"

/clone\_lib="NIH\_BMAP\_M\_S4"

/env\_stage="27-32 days"

/lab\_host="DH10B (Life Technologies)"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker: Site.1: Not I; Site.2: Eco RI; The

NIH\_BMAP\_M\_S4 library is a subtracted library of a series,

ultimately derived from a mixture of individually tagged

normalized libraries from ten regions of the mouse brain

(cerebellum, brain stems, olfactory bulbs, hypothalamus,

cortex, amygdala, basal ganglia, pineal gland, striatum,

hippocampus) after a series of subtractions to reduce the

representation of cDNAs from which ESTs had already been

generated. The following serially subtracted libraries

were generated in this process: NIH\_BMAP\_M\_S4

NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1,

NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library

(NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified

cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and

NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived

was used as a driver in a hybridization with a pool of

the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1

libraries in the form of single-stranded circles. The

remaining single-stranded circles (subtracted library)

was purified by hydroxyapatite column chromatography.

converted to double-stranded circles and electroporated

into DH10B bacteria (lifetechologies) to generate the

NIH\_BMAP\_M\_S4 library. This procedure has been previously

described (Bonaldo, Lennon and Soares, Genome Research

seq_documentation_block:	691 bp	mRNA	linear	EST 06-FEB-2001
LOCUS	BG162342			
DEFINITION	de3ee08.y1 wellcome CRC PRN3 dorsal lip <i>Xenopus laevis</i> cdna clone			
IMAGE:	3473750 5' similar to TR:095432 095432 HYPOTHETICAL 72.5 KD			



```

seq_name: gb_est2:BJ060213

seq_documentation_block:
LOCUS      BJ060213      645 bp      mRNA      linear      EST 11-DEC-2001
DEFINITION BJ060213 NTBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL065c05 5', mRNA sequence.
VERSION    BJ060213.1 GI:17496943
KEYWORDS   EST.
SOURCE     African clawed frog.
ORGANISM   Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae;
            Xenopodinae; Xenopus.
REFERENCE  1 (bases 1 to 645)
AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            Y.
TITLE      Expressed genes in X. laevis embryo
JOURNAL    Unpublished (2001)
COMMENT    Contact: Tadasu Shin-i
            Center for Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshin@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
            1..645
             /organism="Xenopus laevis"
             /db_xref="taxon:8355"
             /clone="XL065c05"
             /clone_lib="NIBB Mochii normalized Xenopus tailbud
             library"
             /tissue_type="whole embryo"
             /dev_stage="stage 25"
BASE COUNT 194 a 147 c 170 g 133 t 1 others
ORIGIN

alignment_scores:
  Quality: 968.00      Length: 214
  Ratio: 4.990         Gaps: 1
  Percent Similarity: 90.654 Percent Identity: 79.907

alignment_block:
US-09-863-824-2 x BJ060213 ..
Align seg 1/1 to: BJ060213 from: 1 to: 645
136 AspSerGluAlaAspLysArgSerGlnHisProGluAsnLysProSerTrpSe 152
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
4 GACTCGGCGCGGAAATGAAATGCAAAAGGAGAACATGCCAGCTGGCC 53

152 rValProSerProAspTrpArgAlaTrpTrpGlnArgSerLeuSerLeuA 169
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
54 GGTTCATCACCTGACTGAGAGAGCTGGTGGCAGAGTTCAGCAACTTGC 103

169 laArgAlaAsnSerGlyAspGlnAspTyr**TyrAspSerThrSerAsp 185
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
104 CCGCTATGATTTACGGGACAGGATTACAATACACAGCACTACTAG 153

186 AspSerAsnPheLeuAsnProProArgGlyTrpAspHisThrAlaProGl 202
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
154 GACAGCAACTTCTTAACCTCTGGGAGGAAGGAATAGACAGTGGCCAG 203

202 yHisArgThrPheGluThrLysAspGlnProGluTyrAspSerThrAspG 219
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
204 CCACCCGAACATTTGATACAAAGGAACAGCCTGAGTATGACTATGTAGTG 253

219 lyGluGlyAspTrpSerLeuTrpSerValCysSerValThrCysGlyAsn 235
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
254 GTGAGGAGATTGGACAGTTGCTCAGTGTGCAGTGTCCACCTGTGTGATGC 303

```

```

236 GlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAlaThrGl 252
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
304 GGTAAACCAAAAGCGCACCTGCTGCTATGCTGTACAGCGACAGA 353

252 uSerArgThrCysAspArgProAsnCysProGlyIleGluAspThrPheA 269
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
354 GTCAGAAGACTTGTGATATGCCAAACTGTCCGGAATCGAAGACACATTC 403

269 rgThrAlaAlaThrGluValSerLeuLeuAlaGlySerGluGluPheAsn 285
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
404 GAACAGCAGCCACTGAAGTCAGCTTCTGCGAGGAATGAGGATTTCAAT 453

286 AlaThrLysLeuPheGluValAspThrAspSerCysGluArgTrpMetSe 302
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
454 GCTACAAAGCTGTTTGTGTGATCTACACAGCTGTGACGCGATGGATGA 503

302 rCysLysSerGluPheLeuLysLysTyrMetHisLysValMetAsnAspL 319
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
504 CTGCAAAAGTGAATTTCTCAAAAGTATATGCACAAAGTGGCTAATGACC 553

319 euProSerCysProCysSerTyrProThr.GluValAlaTyrSerThrAl 335
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
554 TTCCACAGCTGCCCTTGTCTCTACCAACTTGAAGTGGCTTATAGCACCG 603

335 aAspIlePheAspArgIleLysArgLysAspPheArgTrp 348
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
604 TGAATATACGATCGGATCAACGCAAAAGACTTTTCGATGG 643

seq_name: gb_est2:BF302850

seq_documentation_block:
LOCUS      BF302850      968 bp      mRNA      linear      EST 21-NOV-2000
DEFINITION 602032745F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4167658 5',
            mRNA sequence.
ACCESSION  BF302850
VERSION    BF302850.1 GI:11249408
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 968)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabbs@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9458 row: j column: 11
            High quality sequence stop: 726.
FEATURES   Location/Qualifiers
            1..968
             /organism="Mus musculus"
             /strain="FVB/N"
             /db_xref="taxon:10090"
             /clone="IMAGE:4167658"
             /clone_lib="NCI_CGAP_SG2"
             /lab_host="DHI0B (T1 phage-resistant)"
             /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
             NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
             dt. Average insert size 1.3 kb. Constructed by Life
             Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 298 a 242 c 219 g 209 t
ORIGIN

alignment_scores:

```

Quality: 854.00 Length: 156  
Ratio: 5.474 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 99.359

alignment\_block:

US-09-863-824-2 x BF302850

Align seg 1/1 to: BF302850 from: 1 to: 968

```
309 LysLysTyrMetHisLysValMetAsnAspLeuProSerCysProCysSe 325
|||||
1 AAGAAATACATGCACAGGTCATCAACAGCTGCCAGCTGCCCTGCTC 50
325 rTyrProThrGluValAlaLafYrSerThrAlaAspIlePheAspArgIleL 342
|||||
51 CTACCCCTACTGAGTGCGCTACACACAGCTGACATCTTTCACCGCATCA 100
342 ySArgLysAspPheArgTrpLysAspAlaSerGlyProLysGluLysLeu 358
|||||
101 AGCCGAAGGACTTCCGATGGAGATGCTAGTGGGCCCAAGAGAAACTA 150
359 GluIleTyrLysProThrAlaArgTyrCysIleArgSerMetLeuSerLe 375
|||||
151 GAGATCTACAGCCTACTGCTGGTACTGCTATCGCTCTATGCTGCTCCT 200
375 uGluSerThrThrLeuAlaAlaGlnHisCysCysTyrGlyAspAsnMetG 392
|||||
201 GGAGAGTACCACACTGGCTGCCAGCACTGTTGCTATGGTGACACATGC 250
392 lnLeuIleThrArgGlyLysGlyAlaGlyThrProAsnLeuIleSerThr 408
|||||
251 AGCTCATCACCAGGGGCAAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
409 GluPheSerAlaGluLeuHisTyrLysValAspValLeuProTrpIleIl 425
|||||
301 GAGTTCCTGCTGCACTCCATCAAAAGTGGATGTTCTGCCCTGGATAT 350
425 eCysLysGlyAspTrpSerArgTyrAsnGluAlaArgProProAsnAsnG 442
351 CTGCAGGGTGCTGAGGACAGATATATGAGCGCGCGCTCCCAATACG 400
442 lyGlnLysCysThrGluSerProSerAspGluAspTyrIleLysGlnPhe 458
401 GACAGAAGTGACAGAGAGCGCTTCTGATGAGGAGTACATCAACAGTTC 450
459 GlnGluAlaArgGluTyr 464
|||||
451 CAAGAAGCCAGAGAGTAC 468
```

seq\_name: gb\_est2:BE947736

seq\_documentation\_block:

LOCUS BE947736 544 bp mRNA linear EST 03-OCT-2000  
DEFINITION UI-M-BH3-axd-h-03-0-UI.s1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone  
UI-M-BH3-axd-h-03-0-UI 3', mRNA sequence.

ACCESSION BE947736

VERSION BE947736.1 GI:10525495

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 544)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704477

CONTACT: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890  
Email: mEst@mail.nih.gov  
Oligo-dt track not found. Not 1 site shown in beginning of sequence  
is likely internal to the message. cDNA library Preparation: M.B.  
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA  
clones from RESEARCH GENETICS. It should be noted that Bento Soares  
is generating a small number of additional specialized  
non-redundant arrays of BMAP cDNAs whose availability will be  
considered under appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA-No.

FEATURES

source

Location/Qualifiers  
1..544  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BH3-axd-h-03-0-UI"  
/clone\_1lb="NIH\_BMAP\_M\_S4"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT7T3D-pac (Pharmacia) with a modified  
polylinker: Site 1: Not 1; Site 2: Eco RI; The  
NIH\_BMAP\_M\_S4 library is a subtracted library of a series,  
ultimately derived from a mixture of individually tagged  
normalized libraries from ten regions of the mouse brain  
(cerebellum, brain stems, olfactory bulbs, hypothalamus,  
cortex, amygdala, basal ganglia, pineal gland, striatum,  
hippocampus) after a series of subtractions to reduce the  
representation of cDNAs from which ESTs had already been  
generated. The following serially subtracted libraries  
were generated in this process: NIH\_BMAP\_M\_S4,  
NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S1,  
NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library  
(NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified  
cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and  
NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived  
was used as a driver in a hybridization with a pool of  
the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1  
libraries in the form of single-stranded circles. The  
remaining single-stranded circles (subtracted library)  
was purified by hydroxyapatite column chromatography,  
converted to double-stranded circles and electroporated  
into DH10B bacteria (Life Technologies) to generate the  
NIH\_BMAP\_M\_S4 library. This procedure has been previously  
described (Bonaldo, Lennon and Soares, Genome Research  
6:791-806, 1996)  
TAG\_SEQ=None found"

BASE COUNT 146 a 165 c 133 g 100 t

ORIGIN

alignment\_scores:

Quality: 827.50 Length: 183

Ratio: 4.926 Gaps: 1

Percent Similarity: 91.803 Percent Identity: 84.153

alignment\_block:

US-09-863-824-2 x BE947736

Align seg 1/1 to: BE947736 from: 1 to: 544

36 AlaAlaGlyAsnAlaSerGlnAlaGlnLeuGlnAsnLeuAsnValG1 52

|||||

3 GCCCGCGGCAACGTCAGCGGGTCCAGTTGCAGATAACCTCAACCTGGA 52

|||||

52 ySerAspThrThrSerGluThrSerPheSerLeuSerLysGluAlaProA 69

|||||

53 GAGTGACTCCACGTCAGAAACACAGCTTTCCTCTGTCCAAAGAACCCCG 102

|||||

69 rgGluHisLeuAspHisGlnAlaHisGlnProPheProArgProArg 85

|||||

103 AGGAG.....CACCAGGTTGTCCACCAACCTCTCCCGACAGACGA 143

86 PheArgGlnGluThrGlyHisProSerLeuGlnArgAspPheProArgse 102  
 111  
 144 TTCCCTCCAGAGACTGCATCTTCTACTGCAGAGACGGCCCAATC 193  
 102 rPheLeuLeuAspLeuProAsnPhProAspLeuSerLysAlaAspIleA 119  
 111  
 194 CTTTCTCTCTGACCTACCAAACTTTCCAGATCTTTCCAAAGCTGATATCA 243  
 119 snGlyGln\*\*\*ProAsnIleGlnValThrIleGluValValAspGlyPro 135  
 244 ATGGGCAGAAATCCAAACATTCAGGTCCACATAGAGGTGGTTGATGGCCCT 293  
 136 AspSerGluAlaAspLysAspGlnHisProGluAsnLysProSerTrpse 152  
 294 GACTCCGAGCAGAAAGATACAGATCCAGAGATTAACCCAGCTGGTC 343  
 152 rValProSerProAspTrpArgAlaTrpTrpGlnArgSerLeuSerLeuA 169  
 111  
 344 GCTCCAGCTCCCGACTGGCGGCTGGTGGCAGAGTCTTGTCTCTTG 393  
 169 laArgAlaAsnSerGlyAspGlnAspTyr\*\*\*TyrAspSerThrSerAsp 185  
 394 CCAGGACAAATAGTGGGACCCAGGATCAAGATGACAGTACCTCAGAT 443  
 186 AspSerAsnPhLeuAsnProProArgGlyTrpAspHisThrAlaProG 202  
 444 GACGCAACTCTCTCAGTGTCTCTAGAGATGGGACCGCTCCAGCCCA 493  
 202 yHisArgThrPheGluThrLysAspGlnProGluThrLysAspSerThr 218  
 494 ACACCGGAGCTTTTGAACAAAGACGACCCAGAGTATGATTCACAGAC 542

seq\_name: gb\_est2:BI891452

seq\_documentation\_block:  
 LOCUS BI891452 666 bp mRNA linear EST 12-OCT-2001  
 DEFINITION ZF637-3-000707 zebrafish shield stage whole embryo cDNA library  
 MPMP637 Danio rerio cDNA clone MPMP637\_15P5;MPMP637P0515 5',  
 mRNA sequence.  
 ACCESSION BI891452 GI:16098723  
 VERSION BI891452.1  
 KEYWORDS EST.  
 SOURCE zebrafish.  
 ORGANISM Danio rerio  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
 ; Cyprinidae; Danio.  
 REFERENCE 1 (Bases 1 to 666)  
 AUTHORS Clark.M., Anstad,P., Hennig,S., Johnson,S.L. and Lehrach,H.  
 TITLE EST sequencing of a zebrafish shield stage cDNA library normalised  
 by oligonucleotide fingerprinting  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Hennig S  
 Laboratory 123, Dept.Lehrach  
 Max-Planck-Institut fuer Molekulare Genetik  
 Ihnestr.63-73, D-14195 Berlin, Germany  
 Tel: +49 30 8413 1612  
 Fax: +49 30 8413 1380  
 Email: hennig@molgen.mpg.de  
 5' EST sequencing of clones from a zebrafish shield stage library,  
 normalised from 55,000 starting clones by oligonucleotide  
 fingerprinting  
 High quality sequence stop: 666.  
 Location/Qualifiers  
 1. 666  
 /organism="Danio rerio"  
 /db\_xref="taxon:7955"  
 /clone="MPMP637\_15P5;MPMP637P0515"  
 /clone\_lib="Zebrafish shield stage whole embryo cDNA  
 library MPMP637"  
 /tissue\_type="whole embryo"  
 /dev\_stage="shield stage, 6 hrs post-fertilisation"  
 /lab\_host="E.coli, XL1 blue MRF"

FEATURES  
 source

/note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI;  
 oligo-dT-NotI primed, SalI adaptors, directionally cloned,  
 library normalised by oligonucleotide fingerprinting"  
 BASE COUNT 186 a 158 c 193 g 127 t 2 others  
 ORIGIN

alignment\_scores:  
 Quality: 703.50 Length: 218  
 Ratio: 4.043 Gaps: 3  
 Percent Similarity: 79.817 Percent Identity: 61.927

alignment\_block:

US-09-863-824-2 x BI891452 ..

Align seg 1/1 to: BI891452 from: 1 to: 666

123 ProAsnIleGlnValThrIleGluValValAspGlyProAsp...SerGI 138  
 111  
 19 CCCATATACCAAGTACCATAGAGTGGTGACAGCTTGGAGAGTTCTGA 68  
 138 uAlaAspLysAspGlnHisProGluAsnLysProSerTrpSerValPro 155  
 111  
 69 GCCAGAGAAGGGAATCGCTAAAGAAACAAAGCCTGGCTGGCA.....G 112  
 155 erProAspTrpArgAlaTrpTrpGlnArgSerLeuSerLeuAlaArgAla 171  
 111  
 113 CTCCTAACTGGAGAACTGGTGGCAGCGTTTCATCATCTCGTCTCATCT 162  
 172 AsnSerGly.....AspGlnAspTyr\*\*\*TyrAspSe 182  
 163 TCTGTGTCCTCCGAGGGGCTCGAGAACAGAGATTACCCCTACGAGAG 212  
 182 rThrSerAspAspSerAsnPhLeuAsnProProArgGlyTrpAspHist 199  
 111  
 213 CAACACAGAGACAGCAACTTCTCAAGCCACTCGGAGACTGGGAGAGAA 262  
 199 hrAlaProGlyHisArgThrPheGluThrLysAspGlnProGluTyrAsp 215  
 111  
 263 GAGTGAAGAGTGAAGCTGGAGCGGGAAGACAGACCCAGACTGAATACG 312  
 216 SerThrAspGlyGluGlyAspTrpSerLeuTrpSerValCysSerValTh 232  
 111  
 313 TACATACACGGGGAAGCGGACTGGAGTGCTTCCATCCAGCTGATC 362  
 232 rCysGlyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCyst 249  
 111  
 363 CTGTGGAATGGTAACAGAAAGCGGACAGATCGTGGGTTATGCCTGCA 412  
 249 hrAlaThrGluSerArgThrCysAspArgProAsnCysProGlyIleGlu 265  
 111  
 413 CTGCCACAGATCAGGACATGTGACATGCCAGCTGTCCAGGATTGAA 462  
 266 AspThrPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerGI 282  
 111  
 463 GATGATTCAGACGGCAGCAACTGAAGTCACTTGTCTCGCTGCCTGCA 512  
 282 uGluPheAsnAlaThrLysLeuPheGluValAspThrAspSerCysGluA 299  
 111  
 513 AAGACTCAATGCCACAGAGCTGTTTGGNAGTGATACCGACAGCTGTGAGC 562  
 299 rgTrpMetSerCysLysSerGluPheLeuLysLysTyrMetHisLysVal 315  
 111  
 563 CGTGTGATGAATGCAAGAGTGAAGTTTCAAGAA..TACATGAGTAAGTA 611  
 316 MetAsnAspLeuProSerCysProCysSerTyrProThrGluValAlaTy 332  
 111  
 612 GCCACTGATACCCAGCTGCCCTGCTTTTACCACCACCGAGGTGGNCTA 661  
 332 rSer 333  
 111  
 662 CAGC 665

```

seq_name: gb_est1:AA884248

seq_documentation_block:
LOCUS      AA884248                425 bp      mRNA      linear      EST 04-JAN-1999
DEFINITION  am32b09.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1468505 3', mRNA sequence.
ACCESSION  AA884248
VERSION    AA884248
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 425)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgaaps-r@mail.nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert length: 1007 Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 424.
FEATURES   Location/Qualifiers
            source
            1..425
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:1468505"
                /clone_lib="Soares_NFL_T_GBC_S1"
                /lab_host="DH10B"
                /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
                Equal amounts of plasmid DNA from three normalized
                libraries (fetal lung NBH19W, testis NHT, and B-cell
                NCI_CGAP_GCB1) were mixed, and ss circles were made in
                vitro. Following HAP purification, this DNA was used as
                tracer in a subtractive hybridization reaction. The driver
                was PCR-amplified cDNAs from pools of 5,000 clones made
                from the same 3 libraries. The pools consisted of
                I.M.A.G.E. clones 297480-302087, 682632-687239,
                726408-728711, and 729096-731399. Subtraction by Bento
                Soares and M. Fatima Bonaldo."
            BASE COUNT  113 a 140 c 98 g 72 t 2 others
            ORIGIN

alignment_scores:
Quality: 684.50      Length: 141
Ratio: 5.225        Gaps: 1
Percent Similarity: 92.908 Percent Identity: 92.908

alignment_block:
US-09-863-824-2 x AA884248 ..
Align seg 1/1 to: AA884248 from: 1 to: 425

64 SerLysGluAlaProArgGluHisLeuAspHisGlnAlaAlaHisGlnPr 80
|||||
3 TCCAAGAGACGACCAAGGAGCATCTGGACCAACAGGCTGCACACCAACC 52
|||||

80 oPheProArgProArgPheArgGlnGluThrGlyHisProSerLeuGlnA 97
|||||
53 CTTCCTCCAGACCGCATCTCCGACAGAGACGGGGGCACCT.TCATTTGCCAA 101
|||||

97 rgAspPheProArgSerPheLeuLeuAspLeuProAsnPheProAspLeu 113
|||||
102 GAGATTCTCCAGATCTTTCTCTCTGATCTACCAAACTTTCCAGATCTT 151
|||||

114 SerLysAlaAspIleAsnGlyGln***ProAsnIleGlnValThrIleGl 130
|||||
152 TCCAAGGTGATATCAATGGGAGAGATCCAAATATCCAGGTCCACCATAGA 201
|||||

```

```

130 uValValAspGlyProAspSerGluAlaAspLysAspGlnHisProGluA 147
|||||
202 GGTGTCGACGGTCTGACTCTGAAGCAGATAAAGATCAGATCCGGAGA 251
|||||

147 snLysProSerTrpSerValProSerProAspTrpArgAlaTrpTrpGln 163
|||||
252 ATAAGCCAGCTGGTCTGAGTCCATCCCGACTGGCGGCCTGGTGGCAG 301
|||||

164 ArgSerLeuSerLeuAlaGargAlaAsnSerGlyAspGlnAspTyr***Ty 180
|||||
302 AGGTCCCTGTCCTTGGCAGGCAACACAGCGGGACCGAGACTACAAGTA 351
|||||

180 rAspSerThrSerAspAspSerAsnPheLeuAsnProProArg...GlyT 196
|||||
352 CGACAGTACCTCAGAGCAGCAACTCTCAACCCCGCCAGCGTNGNG 401
|||||

196 rpAspHisThrAlaProGlyHis 203
|||||
402 GACCCATACAGCCCGCAGGCCAC 424

seq_name: gb_est2:BI704199

seq_documentation_block:
LOCUS      BI704199                715 bp      mRNA      linear      EST 19-SEP-2001
DEFINITION  zfshtARAMA000029 zebrafish shield stage whole embryo cDNA library
MPMGp637 Danio rerio cDNA clone MPMGP637H1826 5', mRNA sequence.
ACCESSION  BI704199
VERSION    BI704199.1 GI:15680329
KEYWORDS   EST.
SOURCE     zebrafish.
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE  1 (bases 1 to 715)
AUTHORS   Dickmeis,T., Aanstad,P., Clark,M., Fischer,N., Herwig,R., Mourrain
            P., Blader,P., Rosa,F., Lehrach,H. and Strahle,U.
TITLE     Identification of Nodal signaling targets by array analysis of
            induced complex probes
JOURNAL   Dev. Dyn. (2001) In press
COMMENT   Contact: Hennig S
            Laboratory 123, dept. Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1612
            Fax: +49 30 8413 1380
            Email: hennig@molgen.mpg.de
            selected as regulated by Taram-A* signalling by cDNA array analysis
            High quality sequence stop: 715.
FEATURES   Location/Qualifiers
            source
            1..715
                /organism="Danio rerio"
                /db_xref="taxon:7955"
                /clone="MPMGp637H1826"
                /clone_lib="zebrafish shield stage whole embryo cDNA
                library MPMGP637"
                /tissue_type="whole embryo"
                /dev_stage="shield stage, 6 hrs post-fertilisation"
                /lab_host="E.coli, XLI blue MRP"
                /note="vector: pSport1; Site_1: NotI; Site_2: SalI;
                oligo-dT-NotI primed, SalI adaptors, directionally cloned,
                library normalised by oligonucleotide fingerprinting"
            BASE COUNT  194 a 171 c 207 g 142 t 1 others
            ORIGIN

alignment_scores:
Quality: 678.50      Length: 223
Ratio: 3.833        Gaps: 3
Percent Similarity: 79.372 Percent Identity: 62.332

alignment_block:

```

```
US-09-863-824-2 x BI704199 ..
Align seg 1/1 to: BI704199 from: 1 to: 715

94 SerLeuGlnArgAspPheProArgSerPheLeuLeuAspLeuProAsnPh 110
|||||
48 AGCGTCGAGGAGACATCCCTGGAGCTTCTGCTGGACCTGCAGAACTT 97
|||||
110 eProAspLeuSerLysAlaAspIleAsnGlyGln***ProAsnIleGlnV 127
|||||
98 TCCTGATCTCTCCAAAGCAGATATAATGGCGAAGACCCCAATATACAGG 147
|||||
127 alThrIleGluValValAspGlyProAsp...SerGluAlaAspLysAsp 142
|||||
148 TCACCATAGAGTGTGGACAGCTGGAGAGTTCTGAGCCAGAGAGGGA 197
|||||
143 GlnHisProGluAsnLysProSerTrpSerValProSerProAspTrpAr 159
|||||
198 ATGCGTAAGAAACAAAGCCTGGCTGGCGAGCTCTCTAGC.....TGGAG 241
|||||
159 gAlaTrpTrpGlnArgSerLeuSerLeuAlaArgAlaAsnSerGly.... 174
|||||
242 GAACCTGGTGGCAGCTTCATCTCCTCTCATCTCTCTCTCTCTCTCTCT 291
|||||
175 .....AspGlnAspTyr***TyrAspSerThrSerAspAsp 186
|||||
292 CGAAGGGGCGCTGAGGACAGGATTACCCCTAGGAGAGCAACACAGAGAC 341
|||||
187 SerAsnPheLeuAsnProArgGlyTrpAspHisThrAlaProGlyHi 203
|||||
342 AGCAACTTCTCTCAAGCCACTCGGAGACTGGGAGAGAGAGTGAAGAGTGA 391
|||||
203 sargThrPheGluThrLysAspGlnProGluThrAspSerThrAspGlyG 220
|||||
392 AGCTGAGCGGGAAGCAAGCAACCCAGACTGAATACCACTACATAGACGGG 441
|||||
220 luGlyAspTrpSerLeuTrpSerValCysSerValThrCysGlyAsnGly 236
|||||
442 AAGCGGACTGGAGTCTGGTCTCCATGCAGTGTATCTCTGTGGAATGG. 490
|||||
237 AsnGlnLysArgThrArgSerCysGlyTyrAlaCysThrAlaThrGlus 253
|||||
491 AACCAGAAAGCGGACCAAGATCGTGGCGTTATGCTGCACCTGCACACAG 540
|||||
253 rArgThrCysAspArgProAsnCysProGlyIleGluAspThrPheArgT 270
|||||
541 ACCGACATGTGACATCCAGCTGCTCCAGTGTGATGAGATGATTCAGA 590
|||||
270 hrAlaAlaThrGluValSerLeuLeuAlaGlySerGluGluPheAsnAla 286
|||||
591 CGGCAGCAACTGAAGTCAGTTGCTGCTGCTAGCACCTGTGAGTCAATGCA 640
|||||
287 ThrLysLeuPheGluValAspThrAspSerCysGluArgTrpMetSerCy 303
|||||
641 NCAGAGCTGTTGGAGTGTATACCGACAGCTGGGAGCGTTGATGACTGC 690
|||||
303 LysSerGluPheLeuLys 309
|||||
691 AGAGTGAGTGTATAGGAAA 709
|||||

seq_name: gb_est1:AW426875

seq_documentation_block:
LOCUS AW426875 383 bp mRNA linear EST 25-APR-2001
DEFINITION 61687 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW426875
VERSION AW426875.1 GI:6954822
KEYWORDS EST.
SOURCE
  Bos taurus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
  Bovidae; Bovinae; Bos.

REFERENCE
AUTHORS
1 (bases 1 to 383)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perte,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGAGC
Plate: 26 row: L column: 24
Seq primer: ATTTAGGTGACACTATAG.
FEATURES
source
1..383
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT 97 a 104 c 113 g 69 t
ORIGIN
alignment_scores:
Quality: 671.00 Length: 123
Ratio: 5.455 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.374
alignment_block:
US-09-863-824-2 x AW426875 ..
Align seg 1/1 to: AW426875 from: 1 to: 383
233 CysGlyAsnGlyAsnGlnLysArgThrArgSerCysGlyTyrAlaCysTh 249
|||||
14 TGTGGGAATGGCAACCAAGAAACGGACAGGCTTGTGGCTACGCGTGTAC 63
|||||
249 rAlaThrGluSerArgThrCysAspArgProAsnCysProGlyIleGluA 266
|||||
64 TGCAACCGAATCTAGGACGTGTGACCGCTCCAGCTGCCAGGAATTGAAG 113
|||||
266 spThrPheArgThrAlaAlaThrGluValSerLeuLeuAlaGlySerGlu 282
|||||
114 ACACCTTCAGGACAGCTGCCACTGAAGTGAAGTGTCTGCTTCGGGAGAGTGA 163
|||||
283 GluPheAsnAlaThrLysLeuPheGluValAspThrAspSerCysGluAr 299
|||||
164 GAGTTTAAATGCCAACAACTGTTTGAAGTCGACACGGACAGCTGTGAGCG 213
|||||
299 gtrpMetSerCysLysSerGluPheLeuLysLysTyrMethisLysValm 316
|||||
214 CTGGATGAGTTGCAAGAGCGAGTTCTCTAAAGAAATACATGATAGGTGA 263
|||||
316 etAsnAspLeuProSerCysProCysSerTyrProThrGluValAlaTyr 332
|||||
264 TCAACGACCTGCTAGCTAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 313
```



127 a l t h r I l e G l u v a l v a l A s p G l y p r o A s p S e r G l u a l a A s p L y s A s p G l n 143

double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Patima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung Nhlh19w."

BASE COUNT 88 a 109 c 92 g 61 t 6 others  
ORIGIN

alignment\_scores:  
Quality: 594.00 Length: 118  
Ratio: 5.351 Gaps: 0  
Percent Similarity: 94.068 Percent Identity: 93.220

alignment\_block:  
US-09-863-824-2 x W86257 ..  
Align seg 1/1 to: W86257 from: 1 to: 356

284 PheAsnAlaThrLysLeuPheGluValAspThrSerCysGluArgTr 300  
|||||  
2 TTTAATGCCACCAAACTGTTTGAAGTTGACACAGACAGCTGTGACNCTG 51  
300 pMetSerCysLysSerGluPheLeuLysLysTyrMethLysValMetA 317  
|||||  
52 GATGAGCTCCAAAGCAGTTCTTAAAGAGTACATGCACAGGTGATGA 101  
317 snAspLeuProSerCysProCysSerTyrProThrGluValAlaTyrSer 333  
|||||  
102 ATGACCTGCCAGCTGCCCTCTCTACCCACCTGAGTGCCCTACAGC 151  
334 ThrAlaAspIlePheAspArgLysArgLysAspPheArgTrPlysAS 350  
|||||  
152 ACGGCCNACATCTTCGACCTCATCAAGCGCAAGGACTTCGCTGGAAGGA 201  
350 pAlaSerGlyProLysGluLysLeuGluIleTyrLysProThrAlaArgT 367  
|||||  
202 CGCCAGCTGGCCCAAGGAGAACTGGAGATCTACAGCCCACTGCCCGGT 251  
367 yCysIleArgSerMetLeuSerLeuGluSerThrThrLeuAlaAlaGln 383  
|||||  
252 ACTGATCGCTCCATCGTGTCTGAGAGACACACCGCTGCGNCACAG 301  
384 HisCysCysTyrClyAspAsnMetGlnLeuIleThrArgGlyLysGlyAl 400  
|||||  
302 CACTGTGCTACGGCGACAACTGACATGACATCAATCACCAGGNCAGGNGTC 351  
400 agly 401  
:||||  
352 GGCG 355

seq\_name: gb\_estl:AW495535

seq\_documentation\_block:  
LOCUS AW495535 379 bp mRNA linear EST 24-FEB-2000  
DEFINITION UI-M-BH3-avb-e-02-0-UI-s1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone  
UI-M-BH3-avb-e-02-0-UI 3', mRNA sequence.  
ACCESSION AW495535  
VERSION AW495535.1 GI:7065816  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 379)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 9704477

COMMENT

Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mEST@mail.nih.gov  
Oligo-qt track not found, Not I site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone Distribution: Researchers may obtain BMAP cDNA  
clones from RESEARCH GENETICS. It should be noted that Bento Soares  
is generating a small number of additional specialized  
non-redundant arrays of BMAP cDNAs whose availability will be  
considered under appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA=No.

FEATURES

source

Location/Qualifiers

1..379  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BH3-avb-e-02-0-UI"  
/clone\_lib="NIH\_BMAP\_M\_S4"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73b-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
NIH\_BMAP\_M\_S4 library is a subtracted library of a series,  
ultimately derived from a mixture of individually tagged  
normalized libraries from ten regions of the mouse brain  
(cerebellum, brain stems, olfactory bulbs, hypothalamus,  
cortex, amygdala, basal ganglia, pineal gland, striatum,  
hippocampus) after a series of subtractions to reduce the  
representation of cDNAs from which ESTs had already been  
generated. The following serially subtracted libraries  
were generated in this process: NIH\_BMAP\_M\_S4,  
NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1,  
NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library  
(NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified  
cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and  
NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived  
was used as a driver in a hybridization with a pool of  
the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1  
libraries in the form of single-stranded circles. The  
remaining single-stranded circles (subtracted library)  
was purified by hydroxyapatite column chromatography,  
converted to double-stranded circles and electroporated  
into DH10B bacteria (Life Technologies) to generate the  
NIH\_BMAP\_M\_S4 library. This procedure has been previously  
described (Bonaldo, Lennon and Soares, Genome Research  
6:791-806, 1996)  
TAG\_LIB=NIH\_BMAP\_M\_S4  
TAG\_TISSUE=hippocampus  
TAG\_SEQ=TTGGAC"

BASE COUNT 99 a 122 c 89 g 69 t  
ORIGIN

alignment\_scores:  
Quality: 574.50 Length: 128  
Ratio: 4.910 Gaps: 1  
Percent Similarity: 91.406 Percent Identity: 83.594

alignment\_block:  
US-09-863-824-2 x AW495535 ..

Align seg 1/1 to: AW495535 from: 1 to: 379

36 AlaAlaGlyAsnAlaSerGlnAlaGlnLeuGlnAsnLeuAsnValG1 52  
|||||  
3 GCCCGCGCAGCTCAGCGGGTCCAGTTCAGATAACCTCACTGGA 52  
52 ySerAspThrThrSerGluThrSerPheSerLeuSerLysGluAlaProA 69

```

BASE COUNT      140 a      144 c      138 g      127 t
ORIGIN
embryos."

alignment_scores:
  Quality: 564.00      Length: 112
  Ratio: 5.127      Gaps: 1
  Percent Similarity: 98.214      Percent Identity: 93.750

alignment_block:
US-09-863-824-2 x BI535390      ..

Align seg 1/1 to: BI535390 from: 1 to: 549

293 AspThrAspSerCysGluArgTrpMetSerCysLysSerGluPheLeu.L 309
:::|||||:::|||||
214 AACACGGACAGCTGTGATCGTCTGATGAGTTGCAAGAGCGAGTTCTCTATA 263
:::|||||:::|||||
309 ySLysTyrMetHisLysValMetAsnAspLeuProSerCysProCysSer 325
|||||:::|||||:::|||||
264 AGAATAACATGATGATAGGTGATCAACGACCTGCCTAGCTGCCCTGCTCC 313
|||||:::|||||

326 TyrProThrGluValAlaTyrSerThrAlaAspIlePheAspArgIleLys 342
|||||:::|||||
314 TACCCACCGAGCTGGCTTACAGCACAGCTGCATCTTCGACCGCATCAA 363
|||||:::|||||

```

name: qb\_est2:BI535390

documentation\_block:

359 luileTyrIysProThrAlaArgTyrCysIleArgSerMetLeuSerLeu 375  
|||||  
414 AGATCTATAAACCCACAGCCCGGTATTGCATTTCGCTCCATGTGTGCCCTG 463

3000

464 GAGAGACCACCGCTGGCTGCCATCACTGTTGCTACGGCGACACATGCA 513  
392 nLeu1IeThrArgGlyLysGlyAlaGlyThrPro 403

## REFERENCE

```

514  GCTTATCACCAGGGGCAAAAGGGGGGGGACGCC 547
      seq_name: gb_est1:AL545675
      seq documentation block:

```

ETL

DEFINITION	AL545675 LTI_NFL006.PL2 Homo sapiens cDNA clone CS01
ACCESSION	AL545675
VERSION	AL545675.1 GT:12878157

VERSION	ALJ4J073.1	GI:12876137
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

Cnidarian  
Mollusk  
Echinoderm  
Amphibian  
Reptilian  
Bird  
Mammal  
Primate  
Hominid  
Human

Homo sapiens  
Cranialata; Vertebrata;  
Eukaryota; Metazoa; Chordata;  
Mammalia; Eutheria; Primates;  
Catarrhini; Hominoidea;

REFERENCE  
1 (bases 1 to 1032)  
AUTHORS  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE  
Full-length cDNA libraries and normalization

### PCR Primers

**FORWARD: AGGAAACAGCTATGACCAT**

BACKWARD: GTTTCCTCAGTCACGACG

Plate: 126 row: D column: 24

Seq primer: ATTAGGTGACACTATAG.

SOURCE	Location/Qualifiers
1	549

```
source
T. .343
/organism="Bos taurus"
```

/db xref="taxon:9913"

```
/clone_lib="MARC 4BOV"
```

```

/tissue_type="pooled"

```

```

/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI;
was primed with a NotI-oligo(dT) primer.

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2002, 14:46:33 ; Search time 72.05 Seconds  
(without alignments)  
1114.083 Million cell updates/sec

Title: US-09-863-824-2  
Perfect score: 2517  
Sequence: 1 MVRLAAELLGLLLTLH.....CTSPSDEYIKQFQAREY 464

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2160	85.8	391	4	Q9H599 homo sapien
2	926	36.8	658	4	O95432 homo sapien
3	321	12.8	60	4	Q9BQL4 homo sapien
4	137.5	5.5	1121	5	Q966P9 caenorhabdi
5	137.5	5.5	1148	5	O966Q0 caenorhabdi
6	135	5.4	1637	6	Q9XSV8 bos taurus
7	133.5	5.3	461	5	O95S22 drosophila
8	133.5	5.3	685	6	Q9RTS5 bos taurus
9	133.5	5.3	763	5	O9XZD0 drosophila
10	131.5	5.2	660	5	O9XZD0 drosophila
11	127	5.0	765	5	O9U8J9 neospora ca
12	125.5	5.0	873	5	O9V692 drosophila
13	124	4.9	2165	5	O19791 caenorhabdi
14	122.5	4.9	1280	11	O9EPX2 mus musculu
15	120.5	4.8	4123	4	O75851 homo sapien
16	120	4.8	822	4	O9UGT4 homo sapien

→ function

17	119.5	4.7	257	5	Q966K7	Q966K7 caenorhabdi
18	119	4.7	687	5	Q23729	Q23729 cryptospori
19	119	4.7	1444	5	O17591	O17591 caenorhabdi
20	118	4.7	813	5	O9BLJ2	O9BLJ2 ciona intes
21	118	4.7	1062	5	Q19204	Q19204 caenorhabdi
22	118	4.7	5636	4	Q96RW7	Q96RW7 homo sapien
23	117.5	4.7	1235	4	O95428	O95428 homo sapien
24	117	4.6	807	6	O9GLX9	O9GLX9 bos taurus
25	117	4.6	898	5	O76822	O76822 branchiosto
26	117	4.6	2673	4	O96SC3	O96SC3 homo sapien
27	116	4.6	490	5	Q968P5	Q968P5 plasmodium
28	116	4.6	619	4	O9H5Y6	O9H5Y6 homo sapien
29	115	4.6	89	11	Q9R150	Q9R150 cavia porce
30	115	4.6	490	5	O968Q5	O968Q5 plasmodium
31	115	4.6	490	5	O968P8	O968P8 plasmodium
32	115	4.6	490	5	O968P7	O968P7 plasmodium
33	115	4.6	490	5	O968P6	O968P6 plasmodium
34	115	4.6	490	5	O95NX7	O95NX7 plasmodium
35	115	4.6	788	5	O22631	O22631 caenorhabdi
36	115	4.6	1083	5	O9VTT0	O9VTT0 drosophila
37	114	4.5	624	4	O94862	O94862 homo sapien
38	113	4.5	1081	5	O9U631	O9U631 drosophila
39	112.5	4.5	1882	5	O9V693	O9V693 drosophila
40	112	4.4	807	4	O9HCB6	O9HCB6 homo sapien
41	111.5	4.4	614	5	O94674	O94674 plasmodium
42	111.5	4.4	1059	5	P90884	P90884 caenorhabdi
43	110.5	4.4	1785	4	O9Y4V9	O9Y4V9 homo sapien
44	110.5	4.4	1905	5	O9XTP6	O9XTP6 plasmodium
45	110.5	4.4	2413	4	Q9UKJ4	Q9UKJ4 homo sapien

## ALIGNMENTS

RESULT 1  
Q9H599 ID Q9H599 PRELIMINARY; PRT; 391 AA.  
AC Q9H599;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE BA149118.1 (NOVEL PROTEIN) (FRAGMENT).  
GN BA149118.1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wilton S.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; ALI33463; CAC16127.2; -;  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00090; tsp\_1; 1.  
DR SMART; SM00209; TSP1; 1.  
DR PROSITE; PS50092; TSP1; 1.  
FT NON\_TER 1  
SQ SEQUENCE 391 AA; 44556 MW; 3B44913173C95649 CRC64;

Query Match 85.8%; Score 2160; DB 4; Length 391;  
Best Local Similarity 99.5%; Pred. No. 9.8e-201;  
Matches 399; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	74	HQAHAQFPFRFQETGHPSLQDRFPRSFLLDLPNPPDLKADINQXQNIQVTIEWD	133
Db	1	HQAHAQFPFRFQETGHPSLQDRFPRSFLLDLPNPPDLKADINQXQNIQVTIEWD	60
QY	134	GPQSEADKQHPKNSVSPSPDRAWQORSLSARANSQDQDYXYDSTSDSSNFLNPP	193
Db	61	GPQSEADKQHPKNSVSPSPDRAWQORSLSARANSQDQDYXYDSTSDSSNFLNPP	120
QY	194	RGWDHTAPGHRFTETKDKQPEYDSTDGBGWSLWSVCSVTGCGNGKQRTRSCGYACTATES	253

NO function

Db	121	RGWDHTAPGHRTTETDKQPEYDSTDGEGDWSLWSVCSVTGNGNQAKRTRSCGVACTATES	180
QY	254	RTCDRPNCPGIEDTFTTAATEVSVLLAGSEEFNATKLFVDTDCSCEWMSCKSEFLKKYMH	313
Db	181	RTCDRPNCPGIEDTFTTAATEVSVLLAGSEEFNATKLFVDTDCSCEWMSCKSEFLKKYMH	240
QY	314	KVMNDLSPCPCSYPTVEYASTADIFDRKKRDFRWKDGASGPKKLEIYKPTARYCIRSM	373
Db	241	KVMNDLSPCPCSYPTVEYASTADIFDRKKRDFRWKDGASGPKKLEIYKPTARYCIRSM	300
QY	374	SLSESTTLAAOHCCYGDNMQLITRCKGAGTPNLSTEPSAELHYKVDVLPWLICKGWSRY	433
Db	301	SLSESTTLAAOHCCYGDNMQLITRCKGAGTPNLSTEPSAELHYKVDVLPWLICKGWSRY	360
QY	434	NEARPPNNGOKCTESPDEDIYIKOFQAREY	464
Db	361	NEARPPNNGOKCTESPDEDIYIKOFQAREY	391
RESULT	2		
O95432			
ID	O95432	PRELIMINARY;	PRT; 658 AA.
AC	O95432;		
DT	01-MAY-1999	(Tremblrel. 10, Created)	
DT	01-MAY-1999	(Tremblrel. 10, Last sequence update)	
DT	01-JUN-2001	(Tremblrel. 17, Last annotation update)	
DE	HYPOTHETICAL	72.5 KDA PROTEIN.	
OS	Homo sapiens	(Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Multimegabase Sequencing Group;		
RT	*Complete sequence of the gene for serine palmitoyltransferase,		
RT	subunit II found on human chromosome 14.;		
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	UW Multimegabase Sequencing Group;		
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF111168; AAB09622.1;		
DR	InterPro; IPR000884; TSP1.		
DR	Pfam; PF00090; tsp.1; 1.		
DR	SMART; SMO0209; TSP1; 1.		
DR	PROSITE; PS50092; TSP1; 1.		
KW	Hypothetical protein.		
SQ	SEQUENCE 658 AA; 72519 MW; 7E1DF224645B300F CRC64;		
Query Match	36.8%;	Score 926;	DB 4; Length 658;
Best Local Similarity	43.1%;	pred. No. 7.5e-81;	
Matches	196;	Conservative	61; Mismatches 126; Indels 72; Gaps 16;
QY	73	DHQAHPQPRPRFRQETG----	HPSLQDRP-----HPSLQDRP-----PRS----- 102
Db	213	DTQASASDPRLRBEEREARLLPRTLQAEHLHQHCWTVTEPAALT	PGNATPPRTQEVTP 272
QY	103	FLLDLPNPDLSKADGXPNQIVTEWDPGPSEADKD--QHPENKP----	SWSVPS 155
Db	273	LLLELQKLPDLVHATLSTPNFDNQVIKVEDPQAEVSDLLAEPSP	PPQDTLSW-LPA 331
QY	156	PDRAWW-----QRSLSLARANSG-----DQDY-----	XYSTSDSN-----FLNPP 193
Db	332	L-WSFLGWDYKGEEDRAPGEKGEKEDEYDPSIEDGEDQEDKEE	EEQALWFNGTT 390
QY	194	RGWDH--TAPGHRFTETKQPEYDSTDGEGDWSLWSVCSVTGNGNQAKRTRSCGVACTAT	251
Db	391	DNMDOGWLAPGDWVFR--KDSVSD--YEPQKEMSPWSPCSGNGSTGQKQTRPCGCGCAT	447
QY	252	ESRTCDRNPCCGIE--DTFRTAATEVSVLLAGSEEFNATKLFVDTDCSCEWMSCKSEFLK	309

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission."  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006615; AAK68231.1;  
SQ SEQUENCE 1121 AA; 126159 MW; 151B03F47E5BF1BC CRC64;

Query Match 5.5%; Score 137.5; DB 5; Length 1121;  
Best Local Similarity 24.4%; Pred. No. 0.0004;  
Matches 47; Conservative 26; Mismatches 71; Indels 49; Gaps 10;

QY 130 EVVDGPDSEADKQ-----HPENKPSWSVSPDWMQWQSLARANSQDQDYXYS--- 182  
Db 471 ECIDGVSCGASEKILCNQPCPEWS-----QWTAW-----TVCDCRGEEIRLRNRC 521  
QY 183 -TSDSNFLNPPRGWDHTAPGHRFTETKQPEYDSTDGEGDWSLWSVCTCGNGNQRT 241  
Db 522 MNAENNNACDGP-----AQDMSCPYRDCPKWE-----EWGEWADCTTCGGGTQKRL 569  
QY 242 RSC--GYACT--ATESRTCDRPNCPGIEDTFRTAATEVSLLAGSEEFNATKLFVDVDFSC 297  
Db 570 RKCDSGNECSGPGEMRFQIASCPYWGDTWPSGCSVSCGQ-----VC 614  
QY 298 ERWMSK-KSEFLK 309  
Db 615 ERTRKICITDEFLQ 627

RESULT 5

ID Q96600 PRELIMINARY; PRT; 1148 AA.  
AC Q96600;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL PROTEIN C3687.5B.  
GN C3687.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Minx P., Minx M.;  
RT "The sequence of C. elegans cosmid C3687.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission."  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006615; AAK68232.1;  
SQ SEQUENCE 1148 AA; 129319 MW; 3D0714E24A2D202D CRC64;

Query Match 5.5%; Score 137.5; DB 5; Length 1148;  
Best Local Similarity 24.4%; Pred. No. 0.00042;  
Matches 47; Conservative 26; Mismatches 71; Indels 49; Gaps 10;

QY 130 EVVDGPDSEADKQ-----HPENKPSWSVSPDWMQWQSLARANSQDQDYXYS--- 182  
Db 471 ECIDGVSCGASEKILCNQPCPEWS-----QWTAW-----TVCDCRGEEIRLRNRC 521  
QY 183 -TSDSNFLNPPRGWDHTAPGHRFTETKQPEYDSTDGEGDWSLWSVCTCGNGNQRT 241  
Db 522 MNAENNNACDGP-----AQDMSCPYRDCPKWE-----EWGEWADCTTCGGGTQKRL 569  
QY 242 RSC--GYACT--ATESRTCDRPNCPGIEDTFRTAATEVSLLAGSEEFNATKLFVDVDFSC 297  
Db 570 RKCDSGNECSGPGEMRFQIASCPYWGDTWPSGCSVSCGQ-----VC 614  
QY 298 ERWMSK-KSEFLK 309  
Db 615 ERTRKICITDEFLQ 627

RESULT 6

ID Q9XSV8 PRELIMINARY; PRT; 1637 AA.  
AC Q9XSV8;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE SCO-SPONDIN (FRAGMENT).  
GN SCO-SPONDIN.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gobron S., Creveaux I., Didier R., Meinzel R.;  
RT "Characterization of cattle SCO-spondin.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ133488; CAB46239.1;  
DR HSSP; P56682; ICCV.  
DR InterPro; IPR000359; Cys\_knot.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001545; Glyco\_hormone\_beta.  
DR InterPro; IPR002919; TIL.  
DR InterPro; IPR000884; TSPI.  
DR InterPro; IPR001007; VWFC.  
DR Pfam; PF00007; Cys\_knot; 1.  
DR Pfam; PF01826; TIL; 5.  
DR Pfam; PF00090; tsp\_1; 1.  
DR SMART; SM00041; CT; 1.  
DR SMART; SM00001; EGF\_like; 1.  
DR SMART; SM00068; GHB; 1.  
DR SMART; SM00209; TSPI; 1.  
DR SMART; SM00214; VWC; 2.  
DR PROSITE; PS01225; CTCK\_2; 1.  
DR PROSITE; PS00261; GLYCO\_HORMONE\_BETA\_1; UNKNOWN\_2.  
DR PROSITE; PS50092; TSPI; 1.  
DR PROSITE; PS01208; VWFC; UNKNOWN\_1.  
FT NON\_TER 1  
SQ SEQUENCE 1637 AA; 173347 MW; 4C5BAB1DD346C925 CRC64;

Query Match 5.4%; Score 135; DB 6; Length 1637;  
Best Local Similarity 23.4%; Pred. No. 0.0012;  
Matches 59; Conservative 17; Mismatches 70; Indels 106; Gaps 12;

QY 150 SWSVSPDWMQWQSLARANSQDQDYXYSDDSNFLNPPRGWDHTAPGHRFTETK 209  
Db 644 AWS-----RWSAWSPCSRSCGPA--GQSRFRSSTSG-----SW---APECREQSQ 685  
QY 210 DQP-----EYDSTDGEG-----DMSLW 226  
Db 686 SQPCPSQCPPLCQGTFRPRSLGDSWLDQCCQCCCTPEGIICEDAEACAGLAWTPSPW 745  
QY 227 SVCSVTGNGNQKTRSC-----CYACTA--TESRTCDRPNCPGIEDT----- 267

Db 746 SDCPVSQGGNQVTRVCVASAPRGGSPCLGDPDQSORGLWPCPALPDTCSWGPWC 805  
Qy 268 -----FRTAATEVSLAGSEEFNATKLFVDTTSC-----ERWMSCK 304  
Db 806 SRSCGPGSLASASCPCILLAAEPACNATSP-RLDTQACYAGPCLEFCYVSWSSWTRCS 864  
Qy 305 SEEL--KKYMHK 314  
Db 865 CEVLVQQRHQA 876  
RESULT 7  
ID Q95822 PRELIMINARY; PRT; 461 AA.  
AC Q95822.  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DE HL01082P.  
GN FAT-SPONDIN.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=Y, CN BW SP;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celnikier S.  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY060990; AAL28538.1; -  
SQ SEQUENCE 461 AA; 51517 MW; B77D07F41D681B1 CRC64;

Query Match 5.3%; Score 133.5; DB 5; Length 461;  
Best Local Similarity 20.4%; Pred. No. 0.00028;  
Matches 92; Conservative 50; Mismatches 151; Indels 159; Gaps 22;  
Qy 80 PPRPRFRQETGHPSL-----QRDFRSFLDLPNFPDLSKADINGQXPNQVTEV 131  
Db 72 PYNPKSREMTPLAKLYLRREKIVSRNCDDEFLQAL-----QLEV 111  
Qy 132 VGPDSADKQHPENKPSWSPDRAWQORSILSARA-NSGDQDXYDYSTDSDSNFL 190  
Db 112 SD--DAE-EQDTRACRVG-----DYSAWSFCSVSCGKGTMRSRQYLYPAAADQNKCA 162  
Qy 191 NP-----PRGWDHTAPGHRFTFTKQPEYD-----STDGEG----- 221  
Db 163 RLQVAKEMCAIPEACDGA-----QSKDRDDDEGENANSQSLSVSGEGAGLCKTS 216  
Qy 222 DNSLMSVCSVTCGNGNQKRTS-----CGYACTATESRTCDRPN-----PG 263  
Db 217 PWSVNSECSASCIGITRTFVNLGRKRCPH-ITIVKMKMRPCTVEQVELPDQ 275  
Qy 264 IEDTFTATEVSLAGSEEFNATKLFV-----DPTDCERWMSCKSEFLKMYMHKVMNDL 319  
Db 276 CPTSONSDWSPCSCGCGVTRTLRLLENGPDKESCTQME-----LHQQKECV 326  
Qy 320 PSCPCSYPTVEAYSTADIFDRKDRFKWDAAGPKLEIYKPTARYCIRSMLSLETT 379  
Db 327 NPIDCHINAEQA-----KD-----ICVQAPDPGPC-RGYMYAYD 361  
Qy 380 LAAQHC-----CYGDNMQLTTRKGAGTGNLSTERSAELHYKVDVLPWIKGDNWS- 431  
Db 362 PQNHCHYSFTYGCGRNRRNFTDCLNTCNVLRSPYSS-----RVD-QPRACVLSWDSV 416  
Qy 432 -----RYNEARPPNNGKCTE 447

Db 417 WSPCSVSGVGVSESRRYVVTPEFQNGGQPCSK 448  
RESULT 8  
Q9TTS5  
ID Q9TTS5 PRELIMINARY; PRT; 685 AA.  
AC Q9TTS5.  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE SCO-SPONDIN (FRAGMENT).  
GN SCO-SPONDIN.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Theria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SUBCOMMISSURAL ORGAN;  
RA Gobron S., Creveaux I., Monnerie H., Elbitar F.;  
RT "Characterization of cattle SCO-spondin";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ132106; CAB53759.1; -  
DR InterPro; IPR002919; TIL.  
DR InterPro; IPR000884; TSPL.  
DR Pfam; PF01826; TIL; 2.  
DR Pfam; PF00090; tsp\_1; 6.  
DR SMART; SM00209; tsp1; 6.  
DR PROSITE; PS50092; tsp1; 4.  
FT NON\_TER 1 685  
FT NON\_TER 685  
SQ SEQUENCE 685 AA; 71389 MW; A003F880E746D5D3 CRC64;

Query Match 5.3%; Score 133.5; DB 6; Length 685;  
Best Local Similarity 22.6%; Pred. No. 0.00049;  
Matches 78; Conservative 31; Mismatches 137; Indels 99; Gaps 18;  
Qy 146 ENKPSWSPDRAWQORSILSARANSQDQXYDYSTDSDSNFLNPPRGWDHTAGHRT 205  
Db 264 EGQPTWS---PWTWSECSASCPCARRKHFRCTRPGGAPSSMAPPLLLSSVPLCPG 319  
Qy 206 FETKQD---PEYDSTDEGDSLWSVCSVTCGNGKRTSC----- 244  
Db 320 PEAECECLLPEDCRAGGPGWPGWPSWSSCSKSCGGLRSRACDQPPQGLDYCEGPA 379  
Qy 245 -GYACTATESRTCDRPNCGIEDTFTTAATEVSLAGSEEFNATKLFVDTDCSERWMS 303  
Db 380 QGAACQAL---PCPVTNCTALE-----GAE-----YSACGPPCPR--SC 413  
Qy 304 KSEFLKMYMHKVMNDLSPCPCSYPTVEAYSTADIFDRKDRFKWD-ASGPKEK--LEI 360  
Db 414 DD-----LVHCVHCQPCGYCP-PGQVL--SADGTVHVQPGCHSCSCLDLTGERHPPGAQL 465  
Qy 361 YKPT-ARYCIRSMLSLESTL-----AAQHCYGDNNMQLTTRKGAGTGN 404  
Db 466 AKPDGCNYCTCSEGQLTCTDLPCVPGWPCWSEWTACSPQCQ---OTTRSRACSCP- 521  
Qy 405 LISTEFAELHYKVDVLPWIKGDSRYSRNEARPPNNGKCTESP 449  
Db 522 -----APQHGAP-----CPGE---AGEAGAQHRETCASTP 550  
RESULT 9  
ID Q9XZD0 PRELIMINARY; PRT; 763 AA.  
AC Q9XZD0.  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE FAT-SPONDIN PROTEIN.  
GN FAT-SPONDIN OR CG6953.



OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 ON NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Anantadas P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe C.R., Miklos B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Pfeiffer J.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berland B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OREGON-R;  
 RA Darocha S., Baumgartner S.;  
 RT "Fat-Spondin, a Drosophila member of the Spondin family, is highly  
 RT expressed in fat body and hemocytes.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF003804; AAF57910.1;  
 DR EMBL: AF135119; AAD31715.1;  
 DR HSSP: P00974; 1BTL;  
 DR FlyBase: FBgn026721; fat-spondin.  
 DR InterPro: IPR002223; kunitz\_BPTI.  
 DR InterPro: IPR002861; Reeler.  
 DR InterPro: IPR000884; TSPI.  
 DR Pfam: PF00014; Kunitz\_BPTI\_1.  
 DR Pfam: PF02014; Reeler\_1.  
 DR Pfam: PF00090; tsp\_1; 4.  
 DR PRINTS: PRO0759; BASICPTASE.  
 DR SMART: SM00131; KU; 1.  
 DR SMART: SM00209; TSPI; 4.  
 DR PROSITE: PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE: PS0279; BPTI\_KUNITZ\_2; 1.  
 DR PROSITE: PS50092; TSPI; 4.  
 KW Serine protease inhibitor.  
 SQ SEQUENCE 763 AA; 84945 MW; 3292DEDD2CCE4DAB CRC64;

Query Match 5.3%; Score 133.5; DB 5; Length 763;  
 Best Local Similarity 20.4%; Pred. No. 0.00057;  
 Matches 92; Conservative 50; Mismatches 151; Indels 159; Gaps 22;  
 NCBI\_TaxID=7227;  
 QY 80 PFPFRFROETGHPSL-----QRDFPRSFLLDLPLNFPDLSSKADINGQXPNQIVTIEV 131  
 DB 374 PFVNPKSREMTPLAKLYLRREKIVSRNCDEFLQAL-----QLEV 413  
 QY 132 VGPDSADKQHPENKPSWSPDWRAMWORSLSLARA-NSGDODYDYDSTSDSDSNFL 190  
 DB 414 SD--DAE-EQDRAECRVG-----DYSAWSPCSVSGKGRMSRQYLYPAAADONKCA 464  
 QY 191 NP-----PRGMDHTAPCHRTFTKDOPEYD-----STDGEG----- 221  
 DB 465 RQLVAKEMCAVAIPEDAGPA-----QSKRDDDDGGLNLSQSLVSGGEGAGLCKTYS 518  
 QY 222 DWSLMSVSVTCGNGNKRTRS-----COYACTATESRTCDRPNC-----PG 263  
 DB 519 PWSVNSECSASCGIGITMRTFTFVNLGRKRCPH-ITIVKKNKMRPDCTYEQVELPDQ 577  
 QY 264 IEDFTTATEVSLLAGSEEFNATKFEV-----DTSCERWNSCKSEFLKKYMKVNDL 319  
 DB 578 CPTSQMSDWSPCSSTCGRGVITRLLLLLENGDKESCQRM-----LHQOKECV 628  
 QY 320 PSCPCSYPTEVAYSTADIFDRIKRDFRWKDASGPKLEIYKPTARYCIRSMLSLESTT 379  
 DB 629 NPIDCHINAQA-----KD-----ICVQAPDPGCP-RTGYMYAYD 663  
 QY 380 LAAQHC-----CYGDNMQLITRGKAGTNPILISTEFAELHYKVDVLPWIICKGWS- 431  
 DB 664 PQNHCHYSFTYGGCGRNRNFLTNDCLNTCNVLRSPYSS-----RVD-QPRACVLSDWSV 718  
 QY 432 -----RYNEARPPNNGQKCTE 447  
 DB 719 WSPCSVSCGVGVSERRYVYVTEPQNGGQPCSK 750  
 RESULT 10  
 Q23832 PRELIMINARY; PRT; 660 AA.  
 AC Q23832;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE MICRONEMAL TRAP-CI PROTEIN HOMOLOG (FRAGMENT).  
 GN TRAP-CI.  
 OS Cryptosporidium wrairi.  
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriida;  
 OC Cryptosporidiidae; Cryptosporidium.  
 OX NCBI\_TaxID=35083;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GUINEA PIG ISOLATE;  
 RX MEDLINE=98234217; PubMed=9574918;  
 RA Spano F., Putignani L., Naitza S., Puri C., Wright S., Crisanti A.;  
 RT "Molecular cloning and expression analysis of a Cryptosporidium parvum  
 RT gene encoding a new member of the thrombospondin family.";  
 RL Mol. Biochem. Parasitol. 92:147-162(1998).  
 DR EMBL: U42213; AAC48313.1;  
 DR InterPro: IPR000884; TSPI.  
 DR Pfam: PF00090; tsp\_1; 5.  
 DR SMART: SM00209; TSPI; 6.  
 DR PROSITE: PS50092; TSPI; 5.  
 FT NON\_TER 1  
 FT NON\_TER 660  
 SQ SEQUENCE 660 AA; 72812 MW; FFF45490B8812091 CRC64;

Query Match 5.2%; Score 131.5; DB 5; Length 660;  
 Best Local Similarity 22.0%; Pred. No. 0.00073;  
 Matches 69; Conservative 33; Mismatches 116; Indels 95; Gaps 15;

QY 158 WRAMORSLSLARANGDQDYXYDSTSDSNFLNPPRGWDHTAPGHRTFT--KQOPEYD 215  
 Db 352 WSEWSDCSGTS---CGEGNRTIRTEKTP-----PLNGDESTCPELIAKESCNK-VECP 401  
 QY 216 STDGE-GDWSLWSVCSVTCGNGNQKTRSC-GYACT--ATESRTCDRPNCPGIEDTFRTA 271  
 Db 402 NIOCELGWSSWSPCVTCGSGTTSRNRVKGNCTELPTELSKKCNLANC-----451  
 QY 272 ATEVSLLAGSEENATKLFVDVTDSCERWMSCKSE----FLAKYMHKVMNDLPSCPCSYP 327  
 Db 452 -----GDNSASCTAVMSVSE-----WSACEKCDQGLVRRY-----483  
 QY 328 TEVAYSPADIFDRIRKRDFAKMDASGPKELEYKPTARYCTRSMLSLSTTLAAQHOCY 387  
 Db 484 RDFDFSKIGVGVV-----PPGKSEBQNKVREIC-----KDTPTLEEPCTS 525  
 QY 388 GDNQMLITRCKGAGTNPMLISTERSA-----ELHYKVDVLPWLICKGDWSRYNEARP 438  
 Db 526 GVTC-----TPGCKRYTEWSAWSSDCSGSQTDRVTPPEGIIDAICOSSKDRS 575  
 QY 439 PNNGOKCTESPSD 451  
 Db 576 CSRPEGCTETTPD 588  
 RESULT 11  
 Q9U8J9 PRELIMINARY; PRT; 765 AA.  
 AC Q9U8J9  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE THROMBOSPONDIN-RELATED ADHESIVE PROTEIN HOMOLOG.  
 GN Neospora caninum.  
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;  
 OC Neospora.  
 OX NCBI\_TaxID=29176;  
 RN [1]  
 RX MEDLINE=20183852; PubMed=10717300;  
 RA Lovett J.I., Howe D.K., Sibley L.D.;  
 RT "Molecular characterization of a thrombospondin-related anonymous  
 protein homologue in Neospora caninum";  
 RL Mol. Biochem. Parasitol. 107:33-43(2000).  
 DR EMBL: AF061273; AAF01565.1; -  
 DR InterPro: IPR001969; Asp.protease.  
 DR InterPro: IPR000884; TSPI.  
 DR InterPro: IPR002035; VWFA.  
 DR Pfam: PF00090; tsp\_1; 6.  
 DR Pfam: PF00092; vwa; 1.  
 DR PRINTS: PR00453; VWFADOMAIN.  
 DR SMART: SM00209; TSPI; 6.  
 DR SMART: SM00327; VWA; 1.  
 DR PROSITE: PS00141; ASP\_PROTEASE; UNKNOWN\_1.  
 DR PROSITE: PS50092; TSPI; 6.  
 DR PROSITE: PS50234; VWFA; 1.  
 SQ SEQUENCE 765 AA; 82880 MW; 9727838CC1FE8CD4 CRC64;  
 Query Match 5.08; Score 127; DB 5; Length 765;  
 Best Local Similarity 21.88; Pred. No. 0.0025;  
 Matches 60; Conservative 34; Mismatches 81; Indels 100; Gaps 14;  
 QY 135 PDSEADKQHPENKPSWSPV--SPDWRAMWORSLSLARANGSDQDYXYDSTSDSNFLNP 192  
 Db 382 PVEKSSQEQEAFPCVVIPTPEWSAW---SECTATCGGGTRH-----423  
 QY 193 PRGWDHTAPGHRTTE--TKDOPE-----YDSTGE-----GDWSLWSVCSV 231  
 Db 424 -RSRNLGPPGTRSDQNTQEPKPSNPWPFGDLHEQESCNNSPCPINATCGDFGEWSECSV 482  
 QY 232 TCGNGNQKTRT-----SC--GYACTATESRTCDRPNCPGIED-----266

Db 483 SCGGELSORSDRPNWDDQHGKSCMOQYPNGHTKRSNAQPCPVDPEPGDWEWEGCN 542  
 QY 267 -----TFTTAATEVSVLLAGSE---EFNAT-----KLFEVDTDSC-----297  
 Db 543 VTCGGERTRRRRGRSVILPQYGRSIVEQNKSLPENKILLIVETETCSLPCDASCSTFPW 602  
 QY 298 ERWMS---CKSEFLKYYMHKVMN-DLPSCPCSYP 328  
 Db 603 SDWSSCEQCEAGTGTQYRNSAVKFDYRNKPCDFT 637  
 RESULT 12  
 Q9V692 PRELIMINARY; PRT; 873 AA.  
 AC Q9V692  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE CG1739 PROTEIN (GH02025P).  
 GN CG17739.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RX SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfaffler B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reiert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL science 287:2185-2195(2000).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=X, CN BW SP;  
 RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,

RA	Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA	Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA	Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL	Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AE003822; GAF58536.2; -
DR	EMBL; AY058267; AAL13496.1; -
DR	HSPB; P00980; IDTX.
DR	FlyBase; FBgn0033710; CG17739.
DR	InterPro; IPR002223; Kunitz_BPTI.
DR	InterPro; IPR002861; Reeler.
DR	InterPro; IPR000884; TSPI.
DR	Pfam; PF00014; Kunitz_BPTI; 2.
DR	Pfam; PF02014; Reeler; 1.
DR	Pfam; PF00090; tsp_1; 7.
DR	SMART; SM00131; KU; 1.
DR	SMART; SM00209; TSP1; 5.
DR	PROSITE; PS0279; BPTI_KUNITZ_2; 1.
DR	PROSITE; PS0092; TSPI; 5.
DR	Serine protease inhibitor.
KW	SEQUENCE 873 AA; 98281 MW; 62BA97AAFBA42CD1 CRC64;
SQ	

  

Query Match	5.0%; Score 125.5; DB 5; Length 873;
Best Local Similarity	19.4%; Pred. No. 0.0041;
Matches 75; Conservative 31; Mismatches 132; Indels 149; Gaps	

  

Qy	140 DKDQHNPKPSWSVPSPDW-----RAWQRSLSLARNSGDQDYXDSTSD 186
Db	: : : : :           : : : : :   : : : : :
Db	287 DSNHHLISLYSWDPSPDWIGVSGLELCPLNCSSWE-----NKVHNLPWDAGTDS 338
Qy	187 -----SNLNPPRG--WDHTAPGHRTPETKD-----QPEYDS 216
Db	: : : : :           : : : : :   : : : : :
Db	339 GPSYMSADQPVPDVVRRIRKSFPNDPRSFPFDPTGAQMKLATLIHNRRLRYKNCS 398
Qy	217 TDGE-----GDWSLWSCSVTCGNQKRFS-----CGYACTATESPTCD 257
Db	: : : : :           : : : : :   : : : : :
Db	399 SDSEQVPEGCATNSRWMDCTTKCGPGKYRIREFKNPALASRHRCNNA--LREEKNVCV 456
Qy	258 RPNCPGI-EDTFRTAAEVLLAGSSEFNATKLFVDVTDC-----ERWMSCKSEFLKKY 311
Db	: : : : :           : : : : :   : : : : :
Db	457 GHKCAGNFEETAEGGEVEVAPSGSS-----DDQCGLSDHWSSC-----498
Qy	312 MHKVNDLPSCPSYPEVAAYSTADIFDIKRDFRWKDASGPKELEYKPTARYCIIR- 370
Db	: : : : :           : : : : :   : : : : :
Db	499 -----TVTCGTGEM-----TRSRHYLNKKAKKKCKQA 525
Qy	371 SMLSLETTLLAAQHCYCDNNQLITRCKGAGTP-----NLISTEPSAELH 415
Db	: : : : :           : : : : :   : : : : :
Db	526 SRARLHETKICEAMECCGD----IENEGGAGEPEEQAGDGGSAAEKRSIFRNFSYSQ 581
Qy	416 YKVDVLPWIICKGDWSRYNEARPNNNG 442
Db	: : : : :           : : : : :   : : : : :
Db	582 HRDYIPVCGVTPWSDFPSCMGPCGG 608

  

RESULT 13	
ID Q19791	PRELIMINARY; PRT; 2165 AA.
AC Q19791; Q27524;	
DT 01-NOV-1996 (TrEMBLrel. 01, Created)	
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)	
DE F25H8.3 PROTEIN.	
DE F25H8.3	
OS Caenorhabditis elegans.	
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;	
OC Rhabditiidae; Peloderinae; Caenorhabditis.	
OC NCBI_TaxID=6239;	
RN [1]	
RN SEQUENCE FROM N.A.	
RP GaJadsty S.;	
RL Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.	
RN [2]	



Qy	361	-----YKPTARY--CIRSMLSLSTTLAAQHCCYGDNMOL-----ITRGKGACTPN	404
		: :	
Db	3535	LDDCFEADGGFGPSPGPGSCSGGLGTRTRSQR--CVLTMTPLSELPCVCPGPGCGAGN	3592
Qy	405	LISTESAELHYKVDPWPWICK-----GDWSRYNEARPPN-NGQKCTESPSDEDIYIKQF	458
		: :	
Db	3593	CSWTSNA-----PWPECSRCSGVQQRRRLRAYRPPGPGGHWC-----PNILTAY	3636
Qy	459	QEAR	462
		: :	
Db	3637	QERR	3640

Search completed: April 25, 2002, 17:26:33  
Job time: 9600 sec

**THIS PAGE BLANK (USPTO)**



Db 577 SHEYRYLSLREHLAKQRMAGEGNSOVVR-----SLQELLARRTYSGDLLESVDIL- 631  
 QY 423 WLICKGWSRYNEARPPNNGCKTESDDEYIKQFQ 459  
 Db 632 -----RNVDTFRATYVPSADVDQRFQ 655

RESULT 2  
 T00026  
 brain-specific angiogenesis inhibitor 1 - human  
 N:Alternate names: BAI1 protein  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 12-Feb-1999  
 R:Nishimori, H.; Shiratsuchi, T.; Urano, T.; Kimura, Y.; Kiyono, K.; Tatsumi, K.; Yoshida  
 submitted to the EMBL Data Library, June 1997  
 A:Reference number: Z14064  
 A:Accession: T00026  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1584 <NIS>  
 A:Cross-references: EMBL:AB005297; NID:d1175078; PID:d1024528  
 A:Experimental source: brain  
 C:Genetics:  
 A:Gene: GDB:BAI1  
 A:Cross-references: GDB:9838088; OMIM:602682  
 A:Map position: 8q24-8q24  
 C:Superfamily: thrombospondin type 1 repeat homology  
 F:408-462/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 5.5%; Score 137.5; DB 2; Length 1584;  
 Best Local Similarity 21.8%; Pred. No. 0.019;  
 Matches 119; Conservative 50; Mismatches 177; Indels 201; Gaps 30;  
 QY 2 VRLAAELLLLGLLLTLHLITVLRGSAADGPDAAA-----GNASQAQLNNLNVS 53  
 Db 11 WVLAPLILL-LLLGRRAAAGADAGPPECATLVQGFYFSA-----VFP 62  
 QY 54 DTTSETSFSLKEAPREH-LDHOAAHQPPRFRQETGHPSLQDFF-RSFL-----L 105  
 Db 63 ANASRCSMTLRPDPRRYLYMKVAKAPV-----CSGGRVRYQDPSFLESTRYL 115  
 QY 106 DLPNFPD-LSKADINGXENIQVTTEVDGPDSEADKQHPEN---KPSWSVSP--DW- 158  
 Db 116 GVSEFDEVLRCDPSAPLAFQASKQFL-----QMRQPPQHDGLRPRAGPGPTDDES 170  
 QY 159 -----RA-----WQRLSLARAN-----SGDDYXYDSTDD 186  
 Db 171 VEYLVVGNRNPRAAQLCRLWDLACLAGRSRSHPCGIMQTPCACLGGAG----- 221  
 QY 187 SNFLNP-----PRG-----WDHTAPGHR---TFETKQPEYDSTDGEGLWSLWSVTC 233  
 Db 222 ---GPAAGPLAPRGDVCLRDVAGGPENCLSLTDGRGHGATGKWLWSLWGECTRD 277  
 QY 234 GNGNQRTRSC-----GYACTAT---ESRTCDRPNCPGIEDTFRATAAEVSLA----- 279  
 Db 278 GGLQTRTRTCLPAPGVEGGCGEVLEGRQCNREAC---GPAGRTSSRSQSLRSTDAAR 334  
 QY 280 -----GSE-----EFNATKLFVETDSCERWMSCKSEFLKMYMKVMNDLPSCPCSPTEVA 331  
 Db 335 REELGDELQFGPPAQTPQDPAAEWSPWSVCS-----TCG----- 371  
 QY 332 YSTADIFDRKDFRWKDSGPKLELYKPTARYCIRMSLSLE-STTLAAQHCCYGN 390  
 Db 372 -----EGWQTRTECVSSVSTQCSGGLRQRLC---N 401  
 QY 391 MLQITRGKAGTPNLISTEFAELHYKVDVLPWIIIC-----KGDWSRYNEARPPNNGKRC 445  
 Db 402 NSACVPHGAW-----DEWS-----PWSLCSSTCGRFRDTRCTRPQPGGPNP 445  
 QY 446 TSPSDE 452

Db 446 CEGPEKQ 452  
 RESULT 3  
 T00326  
 hypothetical protein KIAA0550 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
 C:Accession: T00326  
 R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,  
 DNA Res. 5, 31-39, 1998  
 A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp  
 A:Reference number: Z14086; MUID:98290545  
 A:Accession: T00326  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-984 <NAG>  
 A:Cross-references: EMBL:AB011122; NID:g3043623; PIDN:BAA25476.1; PID:g3043624  
 A:Experimental source: brain  
 C:Genetics:  
 A:Note: KIAA0550  
 C:Superfamily: thrombospondin type 1 repeat homology  
 F:344-398/Domain: thrombospondin type 1 repeat homology <THR3>  
 QY 62 SLSKEA---PREHL-----DHQAAHQPPRFRQETGHPSLQDFFPSFLD-----LPN 109  
 Db 241 NLTRAKRPPKEEFGMGMDHTIKSQ---RPRSVHEKRVFPQQAADAK-FMAQTGESGVEE 296  
 QY 110 FPDLSKADIN-GOXPNIQVTIEV-----VDGPQSEADKQHPENKPSWSVSPDMDRAW 162  
 Db 297 WSQWSTCVTCGQSQVTRTCVSPYGTCSGLPESRVNCNTALCPVHG-----WEWS 352  
 QY 163 QRLSLARANSGDDYXYDSTSDSNFLNPPRGWDHTAPGHRHTFETKQPEYDSTDGE-G 221  
 Db 353 PWSLCSFTCGRQTRTRSC-----PPQYGRGPCGEPETHHKPCNIALCPVDGQW 404  
 QY 222 DWSLWSVTCGNGNQRTRSC-----GYACTA---TESRTCDRPNCPGIEDTFRAT 273  
 Db 405 EMSSWSQSVTCNSGTQORRQCTAAAHGSGCRGPAESRECYNPEC----- 452  
 QY 274 EVSLLAGSEEFNATKLFVETDSCDCE-----RWMMSCKSEFLK-KYMHKVMNDLPSCP---C 324  
 Db 453 -----TANGQWQWGHWSGCKSCDGGWERRIRTCQGAIVTCQCEGTGEEVRRCSQRC 507  
 QY 325 SYPTVEA---YSTADIFDRIRKDFRWK-----DASGPKEK---LEIY-----KPTARYC 368  
 Db 508 PAPYEICPEDYLMVMVWKRTPDAGDLAFNQCLNATGTTSSRCSLSLHGVAEWFQPSFARC 567  
 QY 369 IRS-----MLESITTLAAQHCCYGNMQLITR 396  
 Db 568 TSNEYRHLQHSIKHEHLAKQRMLAGDGMVSQVTK 600  
 RESULT 4  
 T00028  
 brain-specific angiogenesis inhibitor 3 - human  
 N:Alternate names: BAI3 protein  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000  
 R:Shiratsuchi, T.; Nishimori, H.; Ichise, H.; Nakamura, Y.; Tokino, T.  
 Cyogenet. Cell Genet. 79, 103-108, 1997  
 A:Title: Cloning and characterization of BAI2 and BAI3, novel genes homologous to bra  
 A:Reference number: Z14086; MUID:98194217  
 A:Accession: T00028  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA





F:551-586/Domain: EGF homology <EGF>  
F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 5.18; Score 127.5; DB 2; Length 1170;  
Best Local Similarity 23.3%; Pred. No. 0.08;  
Matches 41; Conservative 17; Mismatches 53; Indels 65; Gaps 8;

QY 151 WSPSPD--WRAWQSRSLARANGDOD--YYXDSTSD-----DSNF 189

DB 373 WPSDADGSPWSPWSSCSATCGNGIQGRSCDNLNRCGSSVQTRTCHIQECKRF 432

QY 190 LNPPRGWHTAPGHRFTET-----KQPEYDSTDGEGD----- 222

DB 433 -KQDGGWHSWSPWSSCSVTCGDGVITRILNCSLSPQMGKPCGEARETRACKKDACPI 491

QY 223 -----WSLWSVCSVTCGNGNQKTRTSC-----GYACT--ATESRTCDRPNCP 262

DB 492 NGGWGWPSPWDICSVTCGGVQRRSLCNPTQPGGKDCVGVDTENQVCNKQDCP 547

# RESULT 7

TSHUPL

C:Species: Homo sapiens (man)  
C:Date: 23-Aug-1987 #sequence\_revision 03-Aug-1995 #text\_change 17-Nov-2000  
C:Accession: A26155; A34274; A30140; A25812; A05172; A42927  
R:Lawler, J.; Hynes, R.O.  
J. Cell Biol. 103, 1635-1648, 1986  
A:Title: The structure of human thrombospondin, an adhesive glycoprotein with multiple  
A:Reference number: A26155; MUID:87057617  
A:Accession: A26155

A:Molecule type: mRNA  
A:Residues: 1-1170 <LAW>  
A:Cross-references: GB:X04665; NID:g37137; PIDN:CAA28370.1; PID:g37138  
A:Note: parts of this sequence, including the amino end of the mature protein, were determined by analysis of the cDNA of the bovine procollagen I N-proteinase.  
R:Lahterty, C.D.; Gierman, T.M.; Dixit, V.M.  
J. Biol. Chem. 264, 11222-11227, 1989  
A:Title: Characterization of the promoter region of the human thrombospondin gene. DNA  
A:Reference number: A34274; MUID:89291870  
A:Accession: A34274

A:Molecule type: DNA  
A:Residues: 1-166 <LAH>  
A:Cross-references: GB:J04835  
R:Hennessy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwein, J. Cell Biol. 108, 729-736, 1989  
A:Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in the 5' noncoding region.  
A:Reference number: A30140; MUID:89139590  
A:Accession: A30140

A:Molecule type: mRNA  
A:Residues: 1-83, 'A', '85-522, 'A', '524-1170 <HEN>  
A:Cross-references: EMBL:X14787; NID:g37464; PIDN:CAA32889.1; PID:g37465  
A:Note: parts of this sequence, including the amino end of the mature protein, were determined by analysis of the cDNA of the bovine procollagen I N-proteinase.  
R:Kobayashi, S.; Eden-McDutchman, F.; Framson, P.; Bornstein, P.  
Biochemistry 25, 8418-8425, 1986  
A:Title: Partial amino acid sequence of human thrombospondin as determined by analysis of the cDNA of the bovine procollagen I N-proteinase.  
A:Reference number: A25812; MUID:87157592  
A:Accession: A25812

A:Molecule type: mRNA  
A:Residues: 1-83, 'A', '85-397 <KOB>  
A:Cross-references: GB:M25631; NID:g538353; PIDN:AAA36741.1; PID:g538354  
R:Dixit, V.M.; Hennessy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.  
Proc. Natl. Acad. Sci. U.S.A. 83, 5449-5453, 1986  
A:Reference number: A05172; MUID:86287276  
A:Accession: A05172

A:Molecule type: mRNA  
A:Residues: 1-83, 'A', '85-374, 'RC' <DIX>  
A:Cross-references: GB:M14326; NID:g340005; PIDN:AAA61237.1; PID:g553801  
A:Note: parts of this sequence, including the amino end of the mature protein, were determined by analysis of the cDNA of the bovine procollagen I N-proteinase.  
R:Sun, X.; Skorstengaard, K.; Mosher, D.F.  
J. Cell Biol. 118, 693-701, 1992  
A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of thrombospondin.  
A:Reference number: A42927; MUID:92348511

A:Accession: A42927  
A:Molecule type: protease  
A:Residues: 987-1003 <SUN>  
A:Note: Cys-992 is shown to have a free sulphydryl  
C:Genetics:  
A:Gene: GDB:THBS1; TSP1; TSP  
A:Cross-references: GDB:120438; OMIM:188060  
A:Map position: 15q15-15q15  
A:Introns: 23/1  
A:Note: the list of introns may be incomplete  
C:Complex: homotrimer, disulfide linked  
C:Function:  
A:Description: participates in cell migration and adhesion, and in platelet aggregation  
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; tri  
C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-1170/Product: thrombospondin 1 #status predicted <MAT>  
F:317-375/Domain: von Willebrand factor type C repeat homology <WVC>  
F:378-429/Domain: thrombospondin type 1 repeat homology <THRI>  
F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>  
F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>  
F:551-586/Domain: EGF homology <EGF1>  
F:650-689/Domain: EGF homology <EGF2>  
F:926-928/Region: cell attachment (R-G-D) motif  
F:171-232/Disulfide bonds: #status predicted  
F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:270,274/Disulfide bonds: interchain #status predicted  
F:610/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted  
F:1051/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 5.08; Score 127; DB 1; Length 1170;

Best Local Similarity 44.4%; Pred. No. 0.087;

Matches 24; Conservative 7; Mismatches 13; Indels 10; Gaps 2;

QY 219 GEGDWSLWSVCSVTCGNGNQKTRTSC-----GYACT--ATESRTCDRPNCP 262

DB 494 GGGPSPWDICSVTCGGVQRRSLCNPTQPGGKDCVGVDTENQVCNKQDCP 547

# RESULT 8

T18517

procollagen N-endopeptidase (EC 3.4.24.14) I - bovine

A:Alternate names: procollagen N-proteinase

C:Species: Bos primigenius taurus (cattle)

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T18517

R:Collige, A.; Nusgens, B.V.; Lapiere, C.M.

submitted to the EMBL Data Library, February 1996

A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.

A:Reference number: Z18941

A:Accession: T18517

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1205 <COL>

A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1

A:Experimental source: skin

C:Genetics:

A:Gene: PC I-NP

C:Function:

A:Description: catalyzes cleavage of the propeptides of type I and II collagens prior

C:Keywords: hydrolase; metalloproteinase

Query Match 4.9%; Score 124; DB 2; Length 1205;

Best Local Similarity 20.9%; Pred. No. 0.16;

Matches 64; Conservative 26; Mismatches 84; Indels 132; Gaps 15;

QY 192 PPRGWDHTAPGHRTE-----TKQPEYDSTDGEGDWSLWS---VCSVTCGNGNQKTRR 242

DB 526 PPLDGTMCAPGKCHKRGHCIIWTLPDILKRD-----GNMGAWSPFGCSSTCTGTGVKFRFR 580

QY 243 SC-----GYACT--ATESRTCDRPNCPGIEDTFRTAATEVSLLAGSEFNATKLFV 292

Db 581 QCDNPHANGRTCSGLAYDFOLCNSQDCPDALADFR----- 617  
Qy 293 DTDSCERWMSKSEFLAKYMHKVMNDLPSCSPYTEVAYSTADIEDRIKRDFRW--- 348  
Db 618 -EEOCRW-----DLIFEHGDQAQHHWLPHE 641  
Qy 349 -KDSAGPKLEIY---KPTARYCIRSMLSLESTLLAAHCYCGDNMQLITRG--KGAGT 402  
Db 642 HRDA---KERCHLYCESKETG---EVVSMKRMVHDGTRCSYKDAFSLCVRGDCRKVGC 693  
Qy 403 PNLISTESFAB-----LHYKVDVLPWLIICKGDSRYNEARPPNNGKQKTESPSDED 453  
Db 694 DGVIGSSKQEDKCGVCGDGNHCKV-----VRGTFSR-----SPKKLG 731  
Qy 454 YIKOFQ 459  
Db 732 YIKMFE 737  
RESULT 9  
T21371  
hypothetical protein F25H8.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T21371; T24896  
R:Gajadaty, S.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z19413  
A:Accession: T21371  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-2165 <W12>  
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3  
A:Experimental source: clone F25H8  
R:Gajadaty, S.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z19949  
A:Accession: T24896  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-2165 <W12>  
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3  
A:Experimental source: clone T13H10  
C:Genetics:  
A:Gene: CESP:F25H8.3  
A:Map position: 4  
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 81

Query Match 4.9%; Score 124; DB 2; Length 2165;  
Best Local Similarity 20.0%; Pred. No. 0.33; Mismatches 129; Indels 164; Gaps 21;  
Matches 83; Conservative 39;  
Qy 141 KDHPENKPSVSPDWR-----AWQSRSLARANSQDQDYXDSTSD-- 186  
Db 1177 KPCHRESCPKYL--GEWSQCSVCEGWSRRYSCVSGNGTEVDSMLCGTASDRPASHQ 1234  
Qy 187 -SNFLNP--RGWDHTA-----PGHRTFTKQDPEYDSTDGE--GD----- 222  
Db 1235 TCNLTGCTCFWRNTDWSACSVCGIGHRETTCTEYIREQSDASFCGDTKMPETSQTCHLL 1294  
Qy 223 ---W--SLASVCSVTCGNGNKRTRSC--GYACTATRESRCDRPNCPGIEDTFTTAATE 274  
Db 1295 PCTSWKPSHSPCSVTCGSGIGTQHSVCTRGSEGTIVDEYFCDRNTRPLRKKT----- 1347  
Qy 275 VSLLAGSEEFNATKLFVEVDYDTSCE-----RWMSCKS-----E 306  
Db 1348 -----CERDTCGPRVLQKLOADVPIRWATGPWTACATCGNGTQR 1390  
Qy 307 FLKYMVKVMNDLPSCPCSPYTEVAYSTADIFDRIKRDFRWKDASGPKLEIYKPTAR 366  
Db 1391 LLKCRDH--VRDLPDEXCNH-----LDKEVSTRNCLRDCS-----YWKMAE 1430

Qy 367 YCIRSMLSLESTLLAAHCYCGDNMQLITRGKAGCTPNLISFEFAEL-----HYKV 418  
Db 1431 W-----EECPATCGTHVQOSRNVTCVSAEDGGRT---ILKDVCDDQKRTPSARNCL 1480  
Qy 419 DVLP-----WII-----CKGWSR-----YNEARPPNNGKQKTE 447  
Db 1481 EPCPKGEBHIGWIIGDWSKCSASCGGWRRRSVCTSSSCDETRKPKMFDKNE 1535  
RESULT 10  
T00355  
hypothetical protein KIAA0688 - human  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
C:Accession: T00355  
R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, DNA Res. 5, 169-176, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. X. The compl A:Reference number: Z14142; MUID:98403880  
A:Accession: T00355  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-837 <I5H>  
A:Cross-references: EMBL:AB014588; NID:g3327189; PIDN:BAA31663.1; PID:g3327190  
A:Experimental source: brain  
C:Genetics:  
A:Gene: KIAA0688  
C:Superfamily: thrombospondin type 1 repeat homology  
F:519-575/Domain: thrombospondin type 1 repeat homology <THR3>  
Query Match 4.8%; Score 122; DB 2; Length 837;  
Best Local Similarity 22.3%; Pred. No. 0.14;  
Matches 61; Conservative 21; Mismatches 88; Indels 104; Gaps 11;  
Qy 104 LLDLPNFPDLSKADINGOXPNIQVTIEVVDGPDSEADKQDQ-----PENKPSWSVSPDWR 159  
Db 424 LLDKPEAP-----LHLPVTF---PGKDYDADQCOLTFGDSRHCPQLPPPCAA 469  
Qy 160 AMQSRSLARANSQDQDYXDSTSDSNFLNPPRGWDHTAPGHRTFTTKDQPEYDST-- 217  
Db 470 LWC-----SGHLN-----GHAMCQTKHSPWADGTPC 495  
Qy 218 -----DGEQWSLWSVCSVTCGNGNKRTRSC----- 244  
Db 496 GPAQACMGRCRLHMDQLQDFNIPQAGWPGMPGWDGCSRTCGGVQVFSRDCRTRPVPRNG 555  
Qy 245 GYACTA--TESRTCDRPNCP--GIEDTFTTAATEVSLLAGSEEFNATKLFVEVDYDTSCEW- 300  
Db 556 GKYPEGRTFRSCNTECDPTGSALTFR---EEQCAAYNHRTDLFKSFPQPMVVPRT 611  
Qy 301 -----MSCKSEFLKKYMHKVMNDLPSCPCS 325  
Db 612 GVAPDQCKLTQCARALGYVYVLEPRVVDGTPCS 645  
RESULT 11  
A42587  
thrombospondin 2 precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Aug-1999  
C:Accession: A42587; A39851  
R:Lahter, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M. J. Biol. Chem. 267, 3274-3281, 1992  
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during ce A:Reference number: A42587; MUID:92147683  
A:Accession: A42587  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-1172 <LAH>  
A:Cross-references: GB:L07803; GB:M87275; NID:g340421; PIDN:AAA53064.1; PID:g567241  
A:Note: sequence extracted from NCBI backbone (NCBIP:81502)

QY 118 INGQXPNIQVTIEVDPDSEADKDQHPENKPSWSVPSP-----DRAWWQORSLARAN 172

R; Bentley, D.

submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of C. elegans cosmid F08C6.  
A:Reference number: Z18440

A:Accession: U15976

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-957 <BEN>

A:Cross-references: EMBL:U29378; NID:g868184; PID:g868185; PIDN:AAA68721.1; CESP:F08C6.1

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:F08C6.1

A:Introns: 23/1; 135/3; 220/3; 285/3; 325/2; 376/3; 461/2; 498/3; 532/3; 627/2; 714/1; 7

Query Match 4.7%; Score 118; DB 2; Length 957;

Best Local Similarity 23.3%; Pred. No. 0.35;

Matches 58; Conservative 27; Mismatches 82; Indels 82; Gaps 14;

QY 221 GDWSLWSVCTCGNGNOKRTSCGYA--CTATESRT--CDRPNCGIEDTFTPTAAATEVS 276

Db 608 GTWSLWTSCTATCGGGYKRNACISITGCEGNEDETEVCSSSECSV----- 655

QY 277 LLAGE-----EFNATKLFVDTDC-----ERWMSCKS-----EFLKYMVKVNDLPS 321

Db 656 LRVGNWSTWTEWNCV-----SCGRSQARYKCLSPHRTLAFCDPGENKVTNEL-- 707

QY 322 CPCSYPTAVYSTADIF-----DRIKKDFRWDA-----SGPKEKLEIYKPTARYC 368

Db 708 -----RITFFKARSYMCVRCNKIKRNTISEKNIEVRSCDNGPCNAIGVGTGWGWS 760

QY 369 IRMSLSLESTTLAQHCYGDNMQLITRGAGTGNLSTFSAELHYKVDYLPWICKG 428

Db 761 TCS-TSCGPGTLVRQTC-----NREPCDGSAAHERRSCNV---ATCQN 799

QY 429 D--WSRYNE 435

Db 800 DGIWSLWNE 808

RESULT 15

A38152

F-spandin - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 17-Nov-2000

A:Accession: A38152

R.Klar A.; Baldassare, M.; Jessell, T.M.

Cell 69, 95-110, 1992

A:Title: F-spandin: a gene expressed at high levels in the floor plate encodes a secreted

A:Reference number: A38152; MUID:92208952

A:Accession: A38152

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-807 <KLA>

A:Cross-references: GB:M88469; NID:g204176; PIDN:AAA41174.1; PID:g204177

A:Experimental source: embryo floor plate

A:Note: sequence extracted from NCBI backbone (NCBIN:90877, NCBI:90878)

C:Superfamily: F-spandin: thrombospondin type 1 repeat homology

F:441-495/Domain: thrombospondin type 1 repeat homology <THR2>

F:500-555/Domain: thrombospondin type 1 repeat homology <THR3>

F:557-611/Domain: thrombospondin type 1 repeat homology <THR4>

F:613-666/Domain: thrombospondin type 1 repeat homology <THR1>

F:667-721/Domain: thrombospondin type 1 repeat homology <THR5>

F:753-807/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match 4.7%; Score 117.5; DB 2; Length 807;

Best Local Similarity 21.0%; Pred. No. 0.31;

Matches 82; Conservative 35; Mismatches 137; Indels 137; Gaps 17;

QY 83 RPRFRQETGHPSLQRDPFRSLDLPNFPD-----LSKADINGQXPNI---QVT 128

Db 368 KPTIPQEKIRPLTSLDHQSPFYD-----PEGGSITQVARVVIERTARKGEQCNIIVPDNDV 423

QY 129 IEVVDGPDSEADKQDHPE-----NKPSW-----SVFSPD- 157

Db 424 DIVADLAPPEKDEDDTPTETCIYSNKPSPWASACSSSTCEKGRMRQRLKAQLDLDSVPCPD 483

QY 158 -----WRAWQORSLARANSQDQDYXYDSTSDSDSNFLNPPR 194

Db 484 QDFQPCMGPGCDEDEGSTCTMSEWITWSPCSVSCGNGMRSRERYVKQFFPEDGVCMLP-- 541

QY 195 GWDHTAPGHRFTETKDOPEYDSTDGEG---DWLSLWSVSVTCGNGNOKRTR-----SC 244

Db 542 -----TEETEKCTVNECSPSSCLVTWGEWDDCSATCGMGKMKRRHRVVKMSPAD 591

QY 245 GYACTA--TESRTCDRPNCPGIEDTFTAAATEVSLIAGSEEFNATKLFEDVTDSCERWMS 302

Db 592 GSMCKAETSQAEEKCMPECHTIP-----CLLSPWSEKSDCSV-----TCGKGMR 635

QY 303 CKSEFLK-----KYMHKVNDLPSCP--CSYPTVAYATAD-----IFDR 340

Db 636 TRQRLMKSLAELGDCNEDLEQAEKCM--LPECPIDCELSSEWSQWSECNKSCKGKGMIRTR 693

QY 341 IKRKDFRWKDAAGPKKEKLEIYKPTARYCIRS 371

Db 694 TIQMEPQFGGACPCP-ETVORKKCRARKCLRS 723

Search completed: April 25, 2002, 17:20:19

Job time: 12041 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 25, 2002, 14:47:28 ; Search time 28.57 seconds  
(without alignments)  
628.837 Million cell updates/sec

Title: US-09-863-824-2  
Perfect score: 2517  
Sequence: 1 MVRLAELLLLGLLLTLH.....CTSPSDEYIKQFEAREY 464

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	143	5.7	1572	1 BAI2_HUMAN	O60241 homo sapien
2	137.5	5.5	1584	1 BAI1_HUMAN	O14514 homo sapien
3	135	5.4	1173	1 TSP1_XENLA	P35448 xenopus lae
4	134.5	5.3	1522	1 BAI3_HUMAN	O60242 homo sapien
5	131	5.2	437	1 PROP_MOUSE	P11680 mus musculus
6	130	5.2	1211	1 AT52_HUMAN	O93450 h adams-2
7	128.5	5.1	1170	1 TSP2_BOVIN	Q95116 bos taurus
8	127.5	5.1	1170	1 TSP1_BOVIN	Q28178 bos taurus
9	127.5	5.1	1170	1 TSP1_MOUSE	P35441 mus musculus
10	127	5.0	1170	1 TSP1_HUMAN	P07996 homo sapien
11	124	4.9	1205	1 AT52_BOVIN	P79331 b adams-2
12	122	4.8	837	1 AT54_HUMAN	O75173 homo sapien
13	120.5	4.8	1172	1 TSP2_MOUSE	Q03350 mus musculus
14	119	4.7	1178	1 TSP2_CHICK	P35440 gallus gall
15	118	4.7	630	1 AT54_RAT	Q96sp7 rattus norv
16	117.5	4.7	807	1 FSP0_RAT	P35446 rattus norv
17	116	4.6	1629	1 AT59_HUMAN	O9p2n4 homo sapien
18	115	4.6	1172	1 TSP2_HUMAN	P35442 homo sapien
19	114	4.5	3562	1 PGCV_CHICK	Q90953 gallus gall
20	111	4.4	469	1 PROP_HUMAN	P27918 homo sapien
21	111	4.4	1201	1 AT53_HUMAN	O15072 homo sapien
22	110.5	4.4	1077	1 SM5A_MOUSE	O62217 mus musculus
23	109	4.3	1093	1 SM5B_MOUSE	O60519 mus musculus
24	107.5	4.3	860	1 AT56_HUMAN	O9ukp5 homo sapien
25	106.5	4.2	803	1 FSP0_XENLA	P35447 xenopus lae
26	106.5	4.2	905	1 AT58_MOUSE	P57110 mus musculus
27	104.5	4.2	559	1 TRAP_PLAFA	P16893 plasmodium
28	104	4.1	470	1 PROP_CAVPO	O64181 cavia porce
29	104	4.1	1074	1 SM5A_HUMAN	O13591 homo sapien
30	102	4.1	592	1 TOR2_CAEEL	P98060 caenorhabdi
31	102	4.1	2026	1 CYAA_YEAST	P08678 saccharomyc
32	99.5	4.0	934	1 C06_HUMAN	P13671 homo sapien
33	99.5	4.0	2333	1 PGCA_CANFA	Q28343 canis famil

RESULT 1

ID	BAI2_HUMAN	STANDARD;	PRT;	1572 AA.
AC	O60241;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Brain-specific angiogenesis inhibitor 2 precursor.			
GN	BAI2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Petal brain;			
RA	Medline=98194217; PubMed=9533023;			
RT	"Cloning and characterization of BAI2 and BAI3, novel genes homologous to brain-specific angiogenesis inhibitor 1 (BAI1).";			
RT	Cytogenet. Cell Genet. 79:103-108(1997).			
CC	- FUNCTION: MIGHT BE INVOLVED IN ANGIOGENESIS INHIBITION.			
CC	- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN BRAIN. ALSO DETECTED IN HEART, THYMUS, SKELETAL MUSCLE, AND DIFFERENT CELL LINES.			
CC	- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.			
CC	- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.			
CC	- SIMILARITY: CONTAINS 1 GPS DOMAIN.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; AB005298; BAA25362.1; -			
DR	MIM; 602683; -			
DR	InterPro: IPR000832; GPCR_secretin.			
DR	InterPro: IPR000203; PKD_cys_rich.			
DR	InterPro: IPR000884; TSP1.			
DR	InterPro: IPR001879; hormn_receptor.			
DR	Pfam: PF00002; 7tm_2; 1.			
DR	Pfam: PF01825; GPS; 1.			
DR	Pfam: PF02793; HRM; 1.			
DR	Pfam: PF00090; tsp_1; 4.			
DR	SMART; SM00303; GPS; 1.			
DR	SMART; SM00008; Hormr; 1.			
DR	SMART; SM00209; TSP1; 4.			
DR	PROSITE; PS00221; GPS; 1.			
DR	PROSITE; PS00649; G_PROTEIN_RECP_F2_1; FALSE_NEG.			
DR	PROSITE; PS00650; G_PROTEIN_RECP_F2_2; FALSE_NEG.			
DR	PROSITE; PS00227; G_PROTEIN_RECP_F2_3; 1.			
DR	PROSITE; PS00261; G_PROTEIN_RECP_F2_4; 1.			
DR	PROSITE; PS50092; TSP1; 4.			

ALIGNMENTS

```
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Repeat.
FT SIGNAL.
FT CHAIN 21 1572
FT DOMAIN 21 924
FT TRANSMEM 925 945
FT TRANSMEM 946 953
FT TRANSMEM 954 974
FT TRANSMEM 975 982
FT TRANSMEM 983 1003
FT TRANSMEM 1004 1024
FT TRANSMEM 1025 1045
FT TRANSMEM 1046 1066
FT TRANSMEM 1067 1087
FT TRANSMEM 1088 1141
FT TRANSMEM 1142 1162
FT TRANSMEM 1163 1168
FT TRANSMEM 1169 1189
FT TRANSMEM 1190 1572
FT DOMAIN 297 351
FT DOMAIN 352 406
FT DOMAIN 407 461
FT DOMAIN 463 517
FT DOMAIN 518 572
FT DOMAIN 573 627
FT DOMAIN 628 682
FT DOMAIN 683 737
FT DOMAIN 738 792
FT DOMAIN 793 847
FT DOMAIN 848 902
FT DOMAIN 903 957
FT DOMAIN 958 1012
FT DOMAIN 1013 1067
FT DOMAIN 1068 1122
FT DOMAIN 1123 1177
FT DOMAIN 1178 1232
FT DOMAIN 1233 1287
FT DOMAIN 1288 1342
FT DOMAIN 1343 1397
FT DOMAIN 1398 1452
FT DOMAIN 1453 1507
FT CARBOHYD 94 94
FT CARBOHYD 179 179
FT CARBOHYD 180 180
FT CARBOHYD 344 344
FT CARBOHYD 425 425
FT CARBOHYD 548 548
FT CARBOHYD 633 633
FT CARBOHYD 855 855
SQ SEQUENCE 1572 AA; 171140 MW; A9775645B77BC285 CRC64;

Query Match 5.7%; Score 143; DB 1; Length 1572;
Best Local Similarity 19.5%; Pred. No. 0.0028;
Matches 89; Conservative 47; Mismatches 161; Indels 160; Gaps 19;

QY 114 SKADINQXPNQIVTVVVDGPDSEADKQHPENKPSW--SVSP-----Dw 158
DB 248 AEADLHSGSSNDLFTTMRVGEPE---EPPKVKQWPRSADEPGLYNAQTGDPAAEW 303
QY 159 RAWQBSLSLARNSGDQDYXDSTSDSNFL-----NPPRGWDHTAPGHTFE---- 207
DB 304 SPMSVCSLT---CGOGLQVTRTSCVSPSYGTLCGSPLRTRPCNNSATCPVHGVEWGS 360
QY 208 -----TKDQPEYDSTDGEG-----DWSLWSVCS 230
DB 361 WSLCSRSGRSGRSRMRTCPVPHOGKACGEPBLOTKCSMAACPVEGOWLEWGPWGPCS 420
QY 231 VTGCGNQKRTSCGYA-----CTA--TESRTCDPNCPIGIEDTFRTAATEVSLLAGSE 382
DB 421 TSCANTQQRKCSVAGPAWATCTCATLDTRECSNLECPATDSKW-----G 467
QY 283 EFNATLKE--VDTDCSRWMSCKSEFLKYYHKVHMNDLPSCPCSYPTVEAYSTAD---- 336
DB 468 PWNWSLCSKTCDTGWRFRMCQATGTQGY-----PCEGTGEVPCSEKRC 516
QY 337 IPDRIRKDF-----RKWDASGPEKELEIYKPTA-----RYCIRS----- 371
DB 517 AFHEMCRDEYVLMVTKWKAAGAEIYKPCPNASGSASRRCLLSAOGVAYWGLPSFARCI 576
QY 372 -----MLSLESTTLAAOHCYGDNMQLITRGKAGTGNLIS--TEFSAEHLHYKVDVLP 422
DB 577 SHEYRYLYLSLREHLAKQBMLAGEGMSQVVR-----SLQELLARTYYSGLDLFSVDIL- 631
```

```
QY 423 WIICKDWSRYNEARPPNNGQKCTESPSDEDIYKQFQ 459
DB 632 -----RNVTDTKRATYVPSADDDVQRFQ 655

RESULT 2
ID BAIL_HUMAN STANDARD; PRT: 1584 AA.
AC O14514;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Brain-specific angiogenesis inhibitor 1 precursor.
GN BAIL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98054121; PubMed=9393972;
RA Nishimori H., Shiratsuchi T., Urano T., Kimura Y., Kiyono K.,
RA Tatsumi K., Yoshida S., Ono M., Kuwano M., Nakamura Y., Tokino T.;
RT "A novel brain-specific p53-target gene, BAIL, containing
RT thrombospondin type 1 repeats inhibits experimental angiogenesis.";
RL Oncogene 15:2145-2150(1997).
RN [2]
RP INTERACTION WITH BAP1.
RX MEDLINE=98321173; PubMed=9647739;
RA Shiratsuchi T., Futamura M., Oda K., Nishimori H., Nakamura Y.,
RA Tokino T.;
RT "Cloning and characterization of BAI-associated protein 1: a PDZ
RT domain-containing protein that interacts with BAIL.";
RL Biochem. Biophys. Res. Commun. 247:597-604(1998).
CC -!- FUNCTION: LIKELY TO BE A POTENT INHIBITOR OF ANGIOGENESIS IN
CC BRAIN AND MAY PLAY A SIGNIFICANT ROLE AS A MEDIATOR OF THE P53
CC SIGNAL IN SUPPRESSION OF GLOBLASTOMA. MAY FUNCTION IN CELL
CC ADHESION AND SIGNAL TRANSDUCTION IN THE BRAIN.
CC -!- SUBUNIT: INTERACTS WITH BAP1.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. LIKELY TO BE
CC CONCENTRATED AT CELL-CELL ADHESION SITES.
CC -!- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN BRAIN. REDUCED OR NO
CC EXPRESSION IS OBSERVED IN SOME GLOBLASTOMA CELL LINES AND CANCER
CC TISSUES.
CC -!- INDUCTION: BY P53.
CC -!- DOMAIN: THE TSP1 REPEATS INHIBIT IN VIVO ANGIOGENESIS IN RAT
CC CORNEA INDUCED BY BFGF.
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; AB005297; BAA23647.1;
DB MIM; 602682;
DB InterPro; IPR000832; GPCR_secretin.
DB InterPro; IPR000203; PKD_cys_rich.
DB InterPro; IPR000884; TSP1.
DB InterPro; IPR001879; hormn_receptor.
DB Pfam; PF00002; 7tm_2; 1.
DB Pfam; PF01825; GPS; 1.
DB Pfam; PF02793; HRM; 1.
DB Pfam; PF00090; tsp.1; 5.
DB SMART; SM00303; GPS; 1.
DB SMART; SM00008; Hormr; 1.
DB SMART; SM00209; TSP1; 5.
```



DR PROSITE; PS50221; GPS; 1.  
DR PROSITE; PS00649; G\_PROTEIN\_RECEP\_F2.1; FALSE\_NEG.  
DR PROSITE; PS00650; G\_PROTEIN\_RECEP\_F2.2; FALSE\_NEG.  
DR PROSITE; PS50227; G\_PROTEIN\_RECEP\_F2.3; 1.  
DR PROSITE; PS50261; G\_PROTEIN\_RECEP\_F2.4; 1.  
DR PROSITE; PS50092; TSP1; 5.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;  
KW Repeat; Cell adhesion.  
FT SIGNAL 1 30  
FT CHAIN 31 1584  
FT DOMAIN 31 948  
FT TRANSMEM 949 969  
FT DOMAIN 970 980  
FT TRANSMEM 981 1001  
FT DOMAIN 1002 1008  
FT TRANSMEM 1009 1029  
FT DOMAIN 1030 1052  
FT TRANSMEM 1053 1073  
FT DOMAIN 1074 1093  
FT TRANSMEM 1094 1114  
FT DOMAIN 1115 1136  
FT TRANSMEM 1137 1157  
FT DOMAIN 1158 1166  
FT TRANSMEM 1167 1187  
FT DOMAIN 1188 1198  
FT TRANSMEM 1199 1218  
FT DOMAIN 1219 1238  
FT TRANSMEM 1239 1258  
FT DOMAIN 1259 1278  
FT TRANSMEM 1279 1298  
FT DOMAIN 1299 1318  
FT TRANSMEM 1319 1338  
FT DOMAIN 1339 1358  
FT TRANSMEM 1359 1378  
FT DOMAIN 1379 1398  
FT TRANSMEM 1399 1418  
FT DOMAIN 1419 1438  
FT TRANSMEM 1439 1458  
FT DOMAIN 1459 1478  
FT TRANSMEM 1479 1498  
FT DOMAIN 1499 1518  
FT TRANSMEM 1519 1538  
FT DOMAIN 1539 1558  
FT TRANSMEM 1559 1578  
FT DOMAIN 1579 1598  
FT TRANSMEM 1599 1618  
FT DOMAIN 1619 1638  
FT TRANSMEM 1639 1658  
FT DOMAIN 1659 1678  
FT TRANSMEM 1679 1698  
FT DOMAIN 1699 1718  
FT TRANSMEM 1719 1738  
FT DOMAIN 1739 1758  
FT TRANSMEM 1759 1778  
FT DOMAIN 1779 1798  
FT TRANSMEM 1799 1818  
FT DOMAIN 1819 1838  
FT TRANSMEM 1839 1858  
FT DOMAIN 1859 1878  
FT TRANSMEM 1879 1898  
FT DOMAIN 1899 1918  
FT TRANSMEM 1919 1938  
FT DOMAIN 1939 1958  
FT TRANSMEM 1959 1978  
FT DOMAIN 1979 1998  
FT TRANSMEM 1999 2018  
FT DOMAIN 2019 2038  
FT TRANSMEM 2039 2058  
FT DOMAIN 2059 2078  
FT TRANSMEM 2079 2098  
FT DOMAIN 2099 2118  
FT TRANSMEM 2119 2138  
FT DOMAIN 2139 2158  
FT TRANSMEM 2159 2178  
FT DOMAIN 2179 2198  
FT TRANSMEM 2199 2218  
FT DOMAIN 2219 2238  
FT TRANSMEM 2239 2258  
FT DOMAIN 2259 2278  
FT TRANSMEM 2279 2298  
FT DOMAIN 2299 2318  
FT TRANSMEM 2319 2338  
FT DOMAIN 2339 2358  
FT TRANSMEM 2359 2378  
FT DOMAIN 2379 2398  
FT TRANSMEM 2399 2418  
FT DOMAIN 2419 2438  
FT TRANSMEM 2439 2458  
FT DOMAIN 2459 2478  
FT TRANSMEM 2479 2498  
FT DOMAIN 2499 2518  
FT TRANSMEM 2519 2538  
FT DOMAIN 2539 2558  
FT TRANSMEM 2559 2578  
FT DOMAIN 2579 2598  
FT TRANSMEM 2599 2618  
FT DOMAIN 2619 2638  
FT TRANSMEM 2639 2658  
FT DOMAIN 2659 2678  
FT TRANSMEM 2679 2698  
FT DOMAIN 2699 2718  
FT TRANSMEM 2719 2738  
FT DOMAIN 2739 2758  
FT TRANSMEM 2759 2778  
FT DOMAIN 2779 2798  
FT TRANSMEM 2799 2818  
FT DOMAIN 2819 2838  
FT TRANSMEM 2839 2858  
FT DOMAIN 2859 2878  
FT TRANSMEM 2879 2898  
FT DOMAIN 2899 2918  
FT TRANSMEM 2919 2938  
FT DOMAIN 2939 2958  
FT TRANSMEM 2959 2978  
FT DOMAIN 2979 2998  
FT TRANSMEM 2999 3018  
FT DOMAIN 3019 3038  
FT TRANSMEM 3039 3058  
FT DOMAIN 3059 3078  
FT TRANSMEM 3079 3098  
FT DOMAIN 3099 3118  
FT TRANSMEM 3119 3138  
FT DOMAIN 3139 3158  
FT TRANSMEM 3159 3178  
FT DOMAIN 3179 3198  
FT TRANSMEM 3199 3218  
FT DOMAIN 3219 3238  
FT TRANSMEM 3239 3258  
FT DOMAIN 3259 3278  
FT TRANSMEM 3279 3298  
FT DOMAIN 3299 3318  
FT TRANSMEM 3319 3338  
FT DOMAIN 3339 3358  
FT TRANSMEM 3359 3378  
FT DOMAIN 3379 3398  
FT TRANSMEM 3399 3418  
FT DOMAIN 3419 3438  
FT TRANSMEM 3439 3458  
FT DOMAIN 3459 3478  
FT TRANSMEM 3479 3498  
FT DOMAIN 3499 3518  
FT TRANSMEM 3519 3538  
FT DOMAIN 3539 3558  
FT TRANSMEM 3559 3578  
FT DOMAIN 3579 3598  
FT TRANSMEM 3599 3618  
FT DOMAIN 3619 3638  
FT TRANSMEM 3639 3658  
FT DOMAIN 3659 3678  
FT TRANSMEM 3679 3698  
FT DOMAIN 3699 3718  
FT TRANSMEM 3719 3738  
FT DOMAIN 3739 3758  
FT TRANSMEM 3759 3778  
FT DOMAIN 3779 3798  
FT TRANSMEM 3799 3818  
FT DOMAIN 3819 3838  
FT TRANSMEM 3839 3858  
FT DOMAIN 3859 3878  
FT TRANSMEM 3879 3898  
FT DOMAIN 3899 3918  
FT TRANSMEM 3919 3938  
FT DOMAIN 3939 3958  
FT TRANSMEM 3959 3978  
FT DOMAIN 3979 3998  
FT TRANSMEM 3999 4018  
FT DOMAIN 4019 4038  
FT TRANSMEM 4039 4058  
FT DOMAIN 4059 4078  
FT TRANSMEM 4079 4098  
FT DOMAIN 4099 4118  
FT TRANSMEM 4119 4138  
FT DOMAIN 4139 4158  
FT TRANSMEM 4159 4178  
FT DOMAIN 4179 4198  
FT TRANSMEM 4199 4218  
FT DOMAIN 4219 4238  
FT TRANSMEM 4239 4258  
FT DOMAIN 4259 4278  
FT TRANSMEM 4279 4298  
FT DOMAIN 4299 4318  
FT TRANSMEM 4319 4338  
FT DOMAIN 4339 4358  
FT TRANSMEM 4359 4378  
FT DOMAIN 4379 4398  
FT TRANSMEM 4399 4418  
FT DOMAIN 4419 4438  
FT TRANSMEM 4439 4458  
FT DOMAIN 4459 4478  
FT TRANSMEM 4479 4498  
FT DOMAIN 4499 4518  
FT TRANSMEM 4519 4538  
FT DOMAIN 4539 4558  
FT TRANSMEM 4559 4578  
FT DOMAIN 4579 4598  
FT TRANSMEM 4599 4618  
FT DOMAIN 4619 4638  
FT TRANSMEM 4639 4658  
FT DOMAIN 4659 4678  
FT TRANSMEM 4679 4698  
FT DOMAIN 4699 4718  
FT TRANSMEM 4719 4738  
FT DOMAIN 4739 4758  
FT TRANSMEM 4759 4778  
FT DOMAIN 4779 4798  
FT TRANSMEM 4799 4818  
FT DOMAIN 4819 4838  
FT TRANSMEM 4839 4858  
FT DOMAIN 4859 4878  
FT TRANSMEM 4879 4898  
FT DOMAIN 4899 4918  
FT TRANSMEM 4919 4938  
FT DOMAIN 4939 4958  
FT TRANSMEM 4959 4978  
FT DOMAIN 4979 4998  
FT TRANSMEM 4999 5018  
FT DOMAIN 5019 5038  
FT TRANSMEM 5039 5058  
FT DOMAIN 5059 5078  
FT TRANSMEM 5079 5098  
FT DOMAIN 5099 5118  
FT TRANSMEM 5119 5138  
FT DOMAIN 5139 5158  
FT TRANSMEM 5159 5178  
FT DOMAIN 5179 5198  
FT TRANSMEM 5199 5218  
FT DOMAIN 5219 5238  
FT TRANSMEM 5239 5258  
FT DOMAIN 5259 5278  
FT TRANSMEM 5279 5298  
FT DOMAIN 5299 5318  
FT TRANSMEM 5319 5338  
FT DOMAIN 5339 5358  
FT TRANSMEM 5359 5378  
FT DOMAIN 5379 5398  
FT TRANSMEM 5399 5418  
FT DOMAIN 5419 5438  
FT TRANSMEM 5439 5458  
FT DOMAIN 5459 5478  
FT TRANSMEM 5479 5498  
FT DOMAIN 5499 5518  
FT TRANSMEM 5519 5538  
FT DOMAIN 5539 5558  
FT TRANSMEM 5559 5578  
FT DOMAIN 5579 5598  
FT TRANSMEM 5599 5618  
FT DOMAIN 5619 5638  
FT TRANSMEM 5639 5658  
FT DOMAIN 5659 5678  
FT TRANSMEM 5679 5698  
FT DOMAIN 5699 5718  
FT TRANSMEM 5719 5738  
FT DOMAIN 5739 5758  
FT TRANSMEM 5759 5778  
FT DOMAIN 5779 5798  
FT TRANSMEM 5799 5818  
FT DOMAIN 5819 5838  
FT TRANSMEM 5839 5858  
FT DOMAIN 5859 5878  
FT TRANSMEM 5879 5898  
FT DOMAIN 5899 5918  
FT TRANSMEM 5919 5938  
FT DOMAIN 5939 5958  
FT TRANSMEM 5959 5978  
FT DOMAIN 5979 5998  
FT TRANSMEM 5999 6018  
FT DOMAIN 6019 6038  
FT TRANSMEM 6039 6058  
FT DOMAIN 6059 6078  
FT TRANSMEM 6079 6098  
FT DOMAIN 6099 6118  
FT TRANSMEM 6119 6138  
FT DOMAIN 6139 6158  
FT TRANSMEM 6159 6178  
FT DOMAIN 6179 6198  
FT TRANSMEM 6199 6218  
FT DOMAIN 6219 6238  
FT TRANSMEM 6239 6258  
FT DOMAIN 6259 6278  
FT TRANSMEM 6279 6298  
FT DOMAIN 6299 6318  
FT TRANSMEM 6319 6338  
FT DOMAIN 6339 6358  
FT TRANSMEM 6359 6378  
FT DOMAIN 6379 6398  
FT TRANSMEM 6399 6418  
FT DOMAIN 6419 6438  
FT TRANSMEM 6439 6458  
FT DOMAIN 6459 6478  
FT TRANSMEM 6479 6498  
FT DOMAIN 6499 6518  
FT TRANSMEM 6519 6538  
FT DOMAIN 6539 6558  
FT TRANSMEM 6559 6578  
FT DOMAIN 6579 6598  
FT TRANSMEM 6599 6618  
FT DOMAIN 6619 6638  
FT TRANSMEM 6639 6658  
FT DOMAIN 6659 6678  
FT TRANSMEM 6679 6698  
FT DOMAIN 6699 6718  
FT TRANSMEM 6719 6738  
FT DOMAIN 6739 6758  
FT TRANSMEM 6759 6778  
FT DOMAIN 6779 6798  
FT TRANSMEM 6799 6818  
FT DOMAIN 6819 6838  
FT TRANSMEM 6839 6858  
FT DOMAIN 6859 6878  
FT TRANSMEM 6879 6898  
FT DOMAIN 6899 6918  
FT TRANSMEM 6919 6938  
FT DOMAIN 6939 6958  
FT TRANSMEM 6959 6978  
FT DOMAIN 6979 6998  
FT TRANSMEM 6999 7018  
FT DOMAIN 7019 7038  
FT TRANSMEM 7039 7058  
FT DOMAIN 7059 7078  
FT TRANSMEM 7079 7098  
FT DOMAIN 7099 7118  
FT TRANSMEM 7119 7138  
FT DOMAIN 7139 7158  
FT TRANSMEM 7159 7178  
FT DOMAIN 7179 7198  
FT TRANSMEM 7199 7218  
FT DOMAIN 7219 7238  
FT TRANSMEM 7239 7258  
FT DOMAIN 7259 7278  
FT TRANSMEM 7279 7298  
FT DOMAIN 7299 7318  
FT TRANSMEM 7319 7338  
FT DOMAIN 7339 7358  
FT TRANSMEM 7359 7378  
FT DOMAIN 7379 7398  
FT TRANSMEM 7399 7418  
FT DOMAIN 7419 7438  
FT TRANSMEM 7439 7458  
FT DOMAIN 7459 7478  
FT TRANSMEM 7479 7498  
FT DOMAIN 7499 7518  
FT TRANSMEM 7519 7538  
FT DOMAIN 7539 7558  
FT TRANSMEM 7559 7578  
FT DOMAIN 7579 7598  
FT TRANSMEM 7599 7618  
FT DOMAIN 7619 7638  
FT TRANSMEM 7639 7658  
FT DOMAIN 7659 7678  
FT TRANSMEM 7679 7698  
FT DOMAIN 7699 7718  
FT TRANSMEM 7719 7738  
FT DOMAIN 7739 7758  
FT TRANSMEM 7759 7778  
FT DOMAIN 7779 7798  
FT TRANSMEM 7799 7818  
FT DOMAIN 7819 7838  
FT TRANSMEM 7839 7858  
FT DOMAIN 7859 7878  
FT TRANSMEM 7879 7898  
FT DOMAIN 7899 7918  
FT TRANSMEM 7919 7938  
FT DOMAIN 7939 7958  
FT TRANSMEM 7959 7978  
FT DOMAIN 7979 7998  
FT TRANSMEM 7999 8018  
FT DOMAIN 8019 8038  
FT TRANSMEM 8039 8058  
FT DOMAIN 8059 8078  
FT TRANSMEM 8079 8098  
FT DOMAIN 8099 8118  
FT TRANSMEM 8119 8138  
FT DOMAIN 8139 8158  
FT TRANSMEM 8159 8178  
FT DOMAIN 8179 8198  
FT TRANSMEM 8199 8218  
FT DOMAIN 8219 8238  
FT TRANSMEM 8239 8258  
FT DOMAIN 8259 8278  
FT TRANSMEM 8279 8298  
FT DOMAIN 8299 8318  
FT TRANSMEM 8319 8338  
FT DOMAIN 8339 8358  
FT TRANSMEM 8359 8378  
FT DOMAIN 8379 8398  
FT TRANSMEM 8399 8418  
FT DOMAIN 8419 8438  
FT TRANSMEM 8439 8458  
FT DOMAIN 8459 8478  
FT TRANSMEM 8479 8498  
FT DOMAIN 8499 8518  
FT TRANSMEM 8519 8538  
FT DOMAIN 8539 8558  
FT TRANSMEM 8559 8578  
FT DOMAIN 8579 8598  
FT TRANSMEM 8599 8618  
FT DOMAIN 8619 8638  
FT TRANSMEM 8639 8658  
FT DOMAIN 8659 8678  
FT TRANSMEM 8679 8698  
FT DOMAIN 8699 8718  
FT TRANSMEM 8719 8738  
FT DOMAIN 8739 8758  
FT TRANSMEM 8759 8778  
FT DOMAIN 8779 8798  
FT TRANSMEM 8799 8818  
FT DOMAIN 8819 8838  
FT TRANSMEM 8839 8858  
FT DOMAIN 8859 8878  
FT TRANSMEM 8879 8898  
FT DOMAIN 8899 8918  
FT TRANSMEM 8919 8938  
FT DOMAIN 8939 8958  
FT TRANSMEM 8959 8978  
FT DOMAIN 8979 8998  
FT TRANSMEM 8999 9018  
FT DOMAIN 9019 9038  
FT TRANSMEM 9039 9058  
FT DOMAIN 9059 9078  
FT TRANSMEM 9079 9098  
FT DOMAIN 9099 9118  
FT TRANSMEM 9119 9138  
FT DOMAIN 9139 9158  
FT TRANSMEM 9159 9178  
FT DOMAIN 9179 9198  
FT TRANSMEM 9199 9218  
FT DOMAIN 9219 9238  
FT TRANSMEM 9239 9258  
FT DOMAIN 9259 9278  
FT TRANSMEM 9279 9298  
FT DOMAIN 9299 9318  
FT TRANSMEM 9319 9338  
FT DOMAIN 9339 9358  
FT TRANSMEM 9359 9378  
FT DOMAIN 9379 9398  
FT TRANSMEM 9399 9418  
FT DOMAIN 9419 9438  
FT TRANSMEM 9439 9458  
FT DOMAIN 9459 9478  
FT TRANSMEM 9479 9498  
FT DOMAIN 9499 9518  
FT TRANSMEM 9519 9538  
FT DOMAIN 9539 9558  
FT TRANSMEM 9559 9578  
FT DOMAIN 9579 9598  
FT TRANSMEM 9599 9618  
FT DOMAIN 9619 9638  
FT TRANSMEM 9639 9658  
FT DOMAIN 9659 9678  
FT TRANSMEM 9679 9698  
FT DOMAIN 9699 9718  
FT TRANSMEM 9719 9738  
FT DOMAIN 9739 9758  
FT TRANSMEM 9759 9778  
FT DOMAIN 9779 9798  
FT TRANSMEM 9799 9818  
FT DOMAIN 9819 9838  
FT TRANSMEM 9839 9858  
FT DOMAIN 9859 9878  
FT TRANSMEM 9879 9898  
FT DOMAIN 9899 9918  
FT TRANSMEM 9919 9938  
FT DOMAIN 9939 9958  
FT TRANSMEM 9959 9978  
FT DOMAIN 9979 9998  
FT TRANSMEM 9999 10018  
FT DOMAIN 10019 10038  
FT TRANSMEM 10039 10058  
FT DOMAIN 10059 10078  
FT TRANSMEM 10079 10098  
FT DOMAIN 10099 10118  
FT TRANSMEM 10119 10138  
FT DOMAIN 10139 10158  
FT TRANSMEM 10159 10178  
FT DOMAIN 10179 10198  
FT TRANSMEM 10199 10218  
FT DOMAIN 10219 10238  
FT TRANSMEM 10239 10258  
FT DOMAIN 10259 10278  
FT TRANSMEM 10279 10298  
FT DOMAIN 10299 10318  
FT TRANSMEM 10319 10338  
FT DOMAIN 10339 10358  
FT TRANSMEM 10359 10378  
FT DOMAIN 10379 10398  
FT TRANSMEM 10399 10418  
FT DOMAIN 10419 10438  
FT TRANSMEM 10439 10458  
FT DOMAIN 10459 10478  
FT TRANSMEM 10479 10498  
FT DOMAIN 10499 10518  
FT TRANSMEM 10519 10538  
FT DOMAIN 10539 10558  
FT TRANSMEM 10559 10578  
FT DOMAIN 10579 10598  
FT TRANSMEM 10599 10618  
FT DOMAIN 10619 10638  
FT TRANSMEM 10639 10658  
FT DOMAIN 10659 10678  
FT TRANSMEM 10679 10698  
FT DOMAIN 10699 10718  
FT TRANSMEM 10719 10738  
FT DOMAIN 10739 10758  
FT TRANSMEM 10759 10778  
FT DOMAIN 10779 10798  
FT TRANSMEM 10799 10818  
FT DOMAIN 10819 10838  
FT TRANSMEM 10839 10858  
FT DOMAIN 10859 10878  
FT TRANSMEM 10879 10898  
FT DOMAIN 10899 10918  
FT TRANSMEM 10919 10938  
FT DOMAIN 10939 10958  
FT TRANSMEM 10959 10978  
FT DOMAIN 10979 10998  
FT TRANSMEM 10999 11018  
FT DOMAIN 11019 11038  
FT TRANSMEM 11039 11058  
FT DOMAIN 11059 11078  
FT TRANSMEM 11079 11098  
FT DOMAIN 11099 11118  
FT TRANSMEM 11119 11138  
FT DOMAIN 11139 11158  
FT TRANSMEM 11159 11178  
FT DOMAIN 11179 11198  
FT TRANSMEM 11199 11218  
FT DOMAIN 11219 11238  
FT TRANSMEM 11239 11258  
FT DOMAIN 11259 11278  
FT TRANSMEM 11279 11298  
FT DOMAIN 11299 11318  
FT TRANSMEM 11319 11338  
FT DOMAIN 11339 11358  
FT TRANSMEM 11359 11378  
FT DOMAIN 11379 11398  
FT TRANSMEM 11399 11418  
FT DOMAIN 11419 11438  
FT TRANSMEM 11439 11458  
FT DOMAIN 11459 11478  
FT TRANSMEM 11479 11498  
FT DOMAIN 11499 11518  
FT TRANSMEM 11519 11538  
FT DOMAIN 11539 11558  
FT TRANSMEM 11559 11578  
FT DOMAIN 11579 11598  
FT TRANSMEM 11599 11618  
FT DOMAIN 11619 11638  
FT TRANSMEM 11639 11658  
FT DOMAIN 11659 11678  
FT TRANSMEM 11679 11698  
FT DOMAIN 11699 11718  
FT TRANSMEM 11719 11738  
FT DOMAIN 11739 11758  
FT TRANSMEM 11759 11778  
FT DOMAIN 11779 11798  
FT TRANSMEM 11799 11818  
FT DOMAIN 11819 11838  
FT TRANSMEM 11839 11858  
FT DOMAIN 11859 11878  
FT TRANSMEM 11879 11898  
FT DOMAIN 11899 11918  
FT TRANSMEM 11919 11938  
FT DOMAIN 11939 11958  
FT TRANSMEM 11959 11978  
FT DOMAIN 11979 11998  
FT TRANSMEM 11999 12018  
FT DOMAIN 12019 12038  
FT TRANSMEM 12039 12058  
FT DOMAIN 12059 12078  
FT TRANSMEM 12079 12098  
FT DOMAIN 12099 12118  
FT TRANSMEM 12119 12138  
FT DOMAIN 12139 12158  
FT TRANSMEM 12159 12178  
FT DOMAIN 12179 12198  
FT TRANSMEM 12199 12218  
FT DOMAIN 12219 12238  
FT TRANSMEM 12239 12258  
FT DOMAIN 12259 12278  
FT TRANSMEM 12279 12298  
FT DOMAIN 12299 12318  
FT TRANSMEM 12319 12338  
FT DOMAIN 12339 12358  
FT TRANSMEM 12359 12378  
FT DOMAIN 12379 12398  
FT TRANSMEM 12399 12418  
FT DOMAIN 12419 12438  
FT TRANSMEM 12439 12458  
FT DOMAIN 12459 12478  
FT TRANSMEM 12479 12498  
FT DOMAIN 12499 12518  
FT TRANSMEM 12519 12538  
FT DOMAIN 12539 12558  
FT TRANSMEM 12559 12578  
FT DOMAIN 12579 12598  
FT TRANSMEM 12599 12618  
FT DOMAIN 12619 12638  
FT TRANSMEM 12639 12658  
FT DOMAIN 12659 12678  
FT TRANSMEM 12679 12698  
FT DOMAIN 12699 12718  
FT TRANSMEM 12719 12738  
FT DOMAIN 12739 12758  
FT TRANSMEM 12759 12778  
FT DOMAIN 12779 12798  
FT TRANSMEM 12799 12818  
FT DOMAIN 12819 12838  
FT TRANSMEM 12839 12858  
FT DOMAIN 12859 12878  
FT TRANSMEM 12879 12898  
FT DOMAIN 12899 12918  
FT TRANSMEM 12919 12938  
FT DOMAIN 12939 12958  
FT TRANSMEM 12959 12978  
FT DOMAIN 12979 12998  
FT TRANSMEM 12999 13018  
FT DOMAIN 13019 13038  
FT TRANSMEM 13039 13058  
FT DOMAIN 13059 13078  
FT TRANSMEM 13079 13098  
FT DOMAIN 13099 13118  
FT TRANSMEM 13119 13138  
FT DOMAIN 13139 13158  
FT TRANSMEM 13159 13178  
FT DOMAIN 13179 13198  
FT TRANSMEM 13199 13218  
FT DOMAIN 13219 13238  
FT TRANSMEM 13239 13258  
FT DOMAIN 13259 13278  
FT TRANSMEM 13279 13298  
FT DOMAIN 13299 13318  
FT TRANSMEM 13319 13338  
FT DOMAIN 13339 13358  
FT TRANSMEM 13359 13378  
FT DOMAIN 13379 13398  
FT TRANSMEM 13399 13418  
FT DOMAIN 13419 13438  
FT TRANSMEM 13439 13458  
FT DOMAIN 13459 13478  
FT TRANSMEM 13479 13498  
FT DOMAIN 13499 13518  
FT TRANSMEM 13519 13538  
FT DOMAIN 13539 13558  
FT TRANSMEM 13559 13578  
FT DOMAIN 13579 13598  
FT TRANSMEM 13599 13618  
FT DOMAIN 13619 13638  
FT TRANSMEM 13639 13658  
FT DOMAIN 13659 13678  
FT TRANSMEM 13679 13698  
FT DOMAIN 13699 13718  
FT TRANSMEM 13719 13738  
FT DOMAIN 13739 13758  
FT TRANSMEM 13759 13778  
FT DOMAIN 13779 13798  
FT TRANSMEM 13799 13818  
FT DOMAIN 13819 13838  
FT TRANSMEM 13839 13858  
FT DOMAIN 13859 13878  
FT TRANSMEM 13879 13898  
FT DOMAIN 13899 13918  
FT TRANSMEM 13919 13938  
FT DOMAIN 13939 13958  
FT TRANSMEM 13959 13978  
FT DOMAIN 13979 13998  
FT TRANSMEM 13999 14018  
FT DOMAIN 14019 14038  
FT TRANSMEM 14039 14058  
FT DOMAIN 14059 14078  
FT TRANSMEM 14079 14098  
FT DOMAIN 14099 14118  
FT TRANSMEM 14119 14138  
FT DOMAIN 14139 14158  
FT TRANSMEM 14159 14178  
FT DOMAIN 14179 14198  
FT TRANSMEM 14199 14218  
FT DOMAIN 14219 14238  
FT TRANSMEM 14239 14258  
FT DOMAIN 14259 14278  
FT TRANSMEM 14279 14298  
FT DOMAIN 14299 14318  
FT TRANSMEM 14319 14338  
FT DOMAIN 14339 14358  
FT TRANSMEM 14359 14378  
FT DOMAIN 14379 14398  
FT TRANSMEM 14399 14418  
FT DOMAIN 14419 14438  
FT TRANSMEM 14439 14458  
FT DOMAIN 14459 14478  
FT TRANSMEM 14479 14498  
FT DOMAIN 14499 14518  
FT TRANSMEM 14519 14538  
FT DOMAIN 14539 14558  
FT TRANSMEM 14559 14578  
FT DOMAIN 14579 14598  
FT TRANSMEM 14599 14618  
FT DOMAIN 14619 14638  
FT TRANSMEM 14639 14658  
FT DOMAIN 14659 14678  
FT TRANSMEM 14679 14698  
FT DOMAIN 14699 14718  
FT TRANSMEM 14719 14738  
FT DOMAIN 14739 14758  
FT TRANSMEM 14759 14778  
FT DOMAIN 14779 14798  
FT TRANSMEM 14799 14818  
FT DOMAIN 14819 14838  
FT TRANSMEM 14839 14858  
FT DOMAIN 14859 14878  
FT TRANSMEM 14879 14898  
FT DOMAIN 14899 14918  
FT TRANSMEM 14919 14938  
FT DOMAIN 14939 14958  
FT TRANSMEM 14959 14978  
FT DOMAIN 14979 14998  
FT TRANSMEM 14999 15018  
FT DOMAIN 15019 15038  
FT TRANSMEM 15039 15058  
FT DOMAIN 15059 15078  
FT TRANSMEM 15079 15098  
FT DOMAIN 15099 15118  
FT TRANSMEM 15119 15138  
FT DOMAIN 15139 15158  
FT TRANSMEM 15159 15178  
FT DOMAIN 15179 15198  
FT TRANSMEM 15199 15218  
FT DOMAIN 15219 15238  
FT TRANSMEM 15239 15258  
FT DOMAIN 15259 15278  
FT TRANSMEM 15279 15298  
FT DOMAIN 15299 15318  
FT TRANSMEM 15319 15338  
FT DOMAIN 15339 15358  
FT TRANSMEM 15359 15378  
FT DOMAIN 15379 15398  
FT TRANSMEM 15399 15418  
FT DOMAIN 15419 15438  
FT TRANSMEM 15439 15458  
FT DOMAIN 15459 15478  
FT TRANSMEM 15479 15498  
FT DOMAIN 15499 15518  
FT TRANSMEM 15519 15538  
FT DOMAIN 15539 15558  
FT TRANSMEM 15559 15578  
FT DOMAIN 15579 15598  
FT TRANSMEM 15599 15618  
FT DOMAIN 15619 15638  
FT TRANSMEM 15639 15658  
FT DOMAIN 15659 15678  
FT TRANSMEM 15679 15698  
FT DOMAIN 15699 15718  
FT TRANSMEM 15719 15738  
FT DOMAIN 15739 15758  
FT TRANSMEM 15759 15778  
FT DOMAIN 15779 15798  
FT TRANSMEM 15799 15818  
FT DOMAIN 15819 15838  
FT TRANSMEM 15839 15858  
FT DOMAIN 15859 15878  
FT TRANSMEM 15879 15898  
FT DOMAIN 15899 15918  
FT TRANSMEM 15919 15938  
FT DOMAIN 15939 15958  
FT TRANSMEM 15959 15978  
FT DOMAIN 15979 15998  
FT TRANSMEM 15999 16018  
FT DOMAIN 16019 16038  
FT TRANSMEM 16039 16058  
FT DOMAIN 16059 16078  
FT TRANSMEM 16079 16098  
FT DOMAIN 16099 16118  
FT TRANSMEM

```
DR SMART; SM00214; VWC; 1.  
DR PROSITE; PS00022; EGF_1; FALSE_NEG.  
DR PROSITE; PS01186; EGF_2; 1.  
DR PROSITE; PS50092; TSP1; 3.  
DR PROSITE; PS01208; VWC; 1.  
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;  
KW EGF-like domain; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 1173  
FT DOMAIN 23 235  
FT DOMAIN 319 376  
FT DOMAIN 382 433  
FT DOMAIN 438 494  
FT DOMAIN 495 546  
FT DOMAIN 550 590  
FT DOMAIN 591 648  
FT DOMAIN 649 693  
FT DOMAIN 726 761  
FT DOMAIN 762 784  
FT DOMAIN 785 820  
FT DOMAIN 821 843  
FT DOMAIN 844 881  
FT DOMAIN 882 917  
FT DOMAIN 918 953  
FT DOMAIN 954 1173  
FT SITE 929 931  
FT DISULFID 554 565  
FT DISULFID 559 575  
FT DISULFID 578 589  
FT DISULFID 595 611  
FT DISULFID 602 620  
FT DISULFID 623 647  
FT DISULFID 653 666  
FT DISULFID 660 679  
FT DISULFID 681 692  
FT CARBOHYD 155 158  
FT CARBOHYD 158 158  
FT CARBOHYD 250 250  
FT CARBOHYD 363 363  
FT CARBOHYD 705 705  
FT CARBOHYD 711 711  
FT CARBOHYD 1070 1070  
SQ SEQUENCE 1173 AA; 130019 MW; A9F036D6516C0F24 CRC64;  
  
Query Match 5.4%; Score 135; DB 1; Length 1173;  
Best Local Similarity 22.3%; Pred. No. 0.0083;  
Matches 59; Conservative 23; Mismatches 88; Indels 94; Gaps 13;  
  
Qy 135 PDSEADKDOHPENKPSVSPDW-----RAWMQSRSLARANGDQDYXDSTS-- 184  
Db 377 PSDSADDWSP-----WS-----DWTGCVTCGHGIDQGRSCDSLNNPCGSSVQTRSCQ 427  
Qy 185 --DSEFLNPPRGWHDYAPGHRFTET-----KDQPEYDSTDGEGD----- 222  
Db 428 IQDCDKRFQDQGWSHWSPSSCVTCGSGQITRILCNLSPVFNOLNCKOCEGEGRENKPC 487  
Qy 223 -----WSLWVSCVTCGNGNQKTRSC-----GYACTA--TESRTCDRP 259  
Db 488 QKDCPINGQWPSLMDTCTVTCGGQMKRRLCNLPKPYQEGKDCIGETDQICNCKQ 547  
Qy 260 NCP-----GIEDTFTAAATEVSLLAGSEE-----FNATKFEVDYDTSRPM 301  
Db 548 DCPIDGCLSNPCFAGVRCV-----SFIDGSKGKCCPPGYRGNGITCKDIDECKEVP 599  
Qy 302 -SKSEFLKYYMKVMDLSPCC 324  
Db 600 DAC---FTLNGVHRCENTPEGYN 620  
  
RESULT 4  
ID BAI3_HUMAN STANDARD; -PRT; 1522 AA.  
  
AC 060242; 060297;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Brain-specific angiogenesis inhibitor 3 precursor.  
GN BAI3 OR KIAA0550.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (LONG ISOFORM).  
RP TISSUE=Fetal brain;  
RX MEDLINE=98194217; PubMed=9533023;  
RA Shiratsuchi T., Nishimori H., Ichise H., Nakamura Y., Tokino T.;  
RT "Cloning and characterization of BAI2 and BAI3, novel genes homologous  
RL to brain-specific angiogenesis inhibitor 1 (BAI1).";  
RL Cytogenet. Cell Genet. 79:103-108(1997).  
RN [2]  
RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
RP TISSUE=Brain;  
RX MEDLINE=98290545; PubMed=9628581;  
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,  
RA Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. IX.  
RT The complete sequences of 100 new cDNA clones from brain which can  
RL code for large proteins in vitro.";  
RL DNA Res. 5:31-39(1998).  
CC -!- FUNCTION: MIGHT BE INVOLVED IN ANGIOGENESIS INHIBITION AND  
CC SUPPRESSION OF GLOBLASTOMA.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN BRAIN. ALSO DETECTED IN  
CC HEART. REDUCED EXPRESSION IS OBSERVED IN SOME GLOBLASTOMA CELL  
CC LINES.  
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.  
CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; AB005299; BAA25363.1; -;  
CC EMBL; AB011122; BAA25476.1; -;  
CC MIM; 602684; -;  
DR InterPro; IPR000859; CUB.  
DR InterPro; IPR000832; GPCR_secretin.  
DR InterPro; IPR000203; PKD_cys_rich.  
DR InterPro; IPR000884; TSP1.  
DR InterPro; IPR001879; hormn_receptor.  
DR Pfam; PF00002; 7tm_2; 1.  
DR Pfam; PF01825; GPS; 1.  
DR Pfam; PF02793; HRM; 1.  
DR Pfam; PF00090; tsp_1; 4.  
DR SMART; SM00303; GPS; 1.  
DR SMART; SM00006; Hormr; 1.  
DR SMART; SM00209; TSP1; 4.  
DR PROSITE; PS01180; CUB; 1.  
DR PROSITE; PS02221; GPS; 1.  
DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.  
DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.  
DR PROSITE; PS00227; G_PROTEIN_RECEP_F2_3; 1.  
DR PROSITE; PS00261; G_PROTEIN_RECEP_F2_4; 1.  
DR PROSITE; PS00092; TSP1; 4.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;  
KW Repeat; Alternative splicing.
```

```
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1522 BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 3.
FT DOMAIN 25 880 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 881 901 1 (POTENTIAL).
FT DOMAIN 902 910 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 911 931 2 (POTENTIAL).
FT DOMAIN 932 939 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 940 960 3 (POTENTIAL).
FT DOMAIN 961 981 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 982 1002 4 (POTENTIAL).
FT DOMAIN 1003 1023 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1024 1044 5 (POTENTIAL).
FT DOMAIN 1045 1098 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1099 1119 6 (POTENTIAL).
FT DOMAIN 1120 1125 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1126 1146 7 (POTENTIAL).
FT DOMAIN 1147 1522 CYTOPLASMIC (POTENTIAL).
FT CUB.
FT DOMAIN 30 159 TSP TYPE-1 1.
FT TRANSMEM 291 344 TSP TYPE-1 2.
FT DOMAIN 345 399 TSP TYPE-1 3.
FT TRANSMEM 400 454 TSP TYPE-1 4.
FT DOMAIN 455 509 GRS.
FT TRANSMEM 816 868 POLY-THR.
FT DOMAIN 942 945 POLY-SER.
FT TRANSMEM 1173 1176 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 51 51 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 82 82 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 337 337 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 625 625 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 779 779 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 812 812 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 828 828 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 937 937 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 643 665 MISSING (IN SHORT ISOFORM).
FT VARSPLIC 990 1007 LPAVVATSVGFTRTGY -> KHIDIPHALKMLNTH
                               (IN SHORT ISOFORM).
FT VARSPLIC 1008 1522 MISSING (IN SHORT ISOFORM).
FT SEQUENCE 1522 AA; 171490 MW; D22D0A5D4BB62502 CRC64;
```

Query Match 5.3%; Score 134.5; DB 1; Length 1522;  
Best Local Similarity 21.4%; Pred. No. 0.013;  
Matches 84; Conservative 48; Mismatches 170; Indels 91; Gaps 21;

```
QY 62 SLKEA---PREHL---DHQAAHQPFRFRQETGHPRLQDRPFRSLD-----LPN 109
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 241 NLTRAKRPKPEFGMGMDHTIKS---RPRSVHEKRVPOEQADAAR-FNAQTGESGVEE 296
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 110 FPLSKADIN-GQXPNIQVTIEV-----VDGPDSEADKDQHPENKPSWSVPSPDRAWW 162
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 297 WSQWSTCSVTGCGSVQVTRTCVSPYTHCSGPLRESRVNNTALCPVHGV---WEEWS 352
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 163 QRLSLARNSGDQDYXYDSTSDSNFLNPPRGWDHTAPGHRFTFKDQPEYDSTGCG-G 221
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 353 PWSLCSFTCGRGORTTRSC-----PPQYGGRCPEGPTHKPKNTALCPVDGQWQ 404
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 222 DWSLWSVCSVTGNGNOKRTRSC-----GYACTA---TESRTCDRPNCPGIEDTFRTAAT 273
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 405 ENWSWOCSTCSNGTOORSQCTAAAHGSGECRGPAWESRECVNPEC----- 452
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 274 EVSLLAGSEEFNATKLFVDYDTSCE-----RWSCKSEFLK-KYMHKVMNDLPSCP---C 324
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 453 -----TANGQWNGHWSGCKSCDGGWERRIRTCQCAVITGQCEGTGEVRRCSQRQ 507
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 325 SYTEVA---YSTADIFDRKDKDFRWK-----DASGPKK---LEIY-----KPTARYC 368
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 508 PAPYEICPEDYLSMWNKRTKTPAGDLAFNQCPLNATGTTSRRCSLSLHGVAFWEPQSPARC 567
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
QY 369 IRS-----MLSESTTLAAQHCYCCYGDNMQLITR 396
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 568 ISNEYRHLQHSIREHLAKGQMLAGCMQSVTK 600
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
ID PROP_MOUSE STANDARD; PRT; 437 AA.
AC P11680;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 13-JUL-1999 (Rel. 38, Last annotation update)
DE Properdin (Fragment).
GN PFC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Macrophage;
RX MEDLINE=88318954; PubMed=3045564;
RA Goundis A., Reid K.B.M.;
RT "Properdin, the terminal complement components, thrombospondin and
   the circumsporozoite protein of malaria parasites contain similar
   sequence motifs.";
RL Nature 335:82-85(1988).
CC -1- FUNCTION: A POSITIVE REGULATOR OF THE ALTERNATE PATHWAY OF
   COMPLEMENT. IT BINDS TO AND STABILIZES THE C3-AND C5-CONVERTASE
   ENZYME COMPLEXES.
CC -1- SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X12905; CAA31389.1; -.
CC PIR; S05478; S05478.
CC MGD; MGI:97545; PFC.
CC InterPro; IPR000884; TSP1.
CC Pfam; PF000090; tsp.1; 6.
CC SMART; SM00209; TSP1; 5.
CC DR PROSITE; PS00092; TSP1; 6.
CC Complement alternate pathway; Glycoprotein; Repeat.
CC NON_TER 1
CC DOMAIN 46 103 TSP TYPE-1 1.
CC DOMAIN 104 160 TSP TYPE-1 2.
CC DOMAIN 161 224 TSP TYPE-1 3.
CC DOMAIN 225 282 TSP TYPE-1 4.
CC DOMAIN 283 345 TSP TYPE-1 5.
CC DOMAIN 346 405 TSP TYPE-1 6.
CC CARBOHYD 396 396 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SEQUENCE 437 AA; 47538 MW; 2B8DBCE22B3B78BE CRC64;

Query Match 5.2%; Score 131; DB 1; Length 437;
Best Local Similarity 26.5%; Pred. No. 0.005;
Matches 53; Conservative 16; Mismatches 79; Indels 52; Gaps 11;

QY 155 SPDWRAWQWSLSLARNSGDQDYXYDSTSDSNFLNPPRGW---DITAPGHRTFE---T 208
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 46 SPQWSASLWGPCSVTCSGSQLRRRCVG-----RGQCSENVAPGTLSEWLOAC 96
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 209 KDQPEYDSTGEGDWSLWSVCSVTGNGNOKRTR-----SCGYAC--TATESRTCD-R 258
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 97 EDQPCCEMGWSWGWPGWPCSVTCSGKTQIRQVCONPAPKCGCHGCGEAAQSQACDTQ 156
   :|:|:| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 259 PNCPIGIEDTFRTAAT-----EVSLLAGSEEFNATKLFVDYDTSDCER-----WMSC 303
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Db 157 KTCP-----THGAWASGWPSPRSGSLGGAQBPKEKTR-----SRSCAPAPSHQPPGKPC 207

QY 304 KSEFLKKYHKHVNLDLPSCP 323

Db 208 SG----PAYEHKACSGLPSCP 224

RESULT 6

AT2S2\_HUMAN

ID AT2S2\_HUMAN STANDARD; PRT; 1211 AA.

AC O95450;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and

DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS2)

DE (procollagen 1/II amino-propeptide processing enzyme) (Procollagen I

DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI)

DE (procollagen 1/II amino-propeptide processing enzyme).

GN ADAMTS2 OR PCINP OR PCPNI.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI).

RC TISSUE=Skin;

RC MEDLINE=99347993; PubMed=10417273;

RA Colige A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,

RA Wertelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,

RA Byers P.H., Lapierre C.M., Prockop D.J., Nussgens B.V.;

RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis

RT are caused by mutations in the procollagen I N-proteinase gene.";

RL Am. J. Hum. Genet. 65:308-317(1999).

CC -I- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR

CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO

CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDENT OF ITS ROLE IN

CC COLLAGEN BIOSYNTHESIS.

CC -I- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain

CC alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains

CC at Ala-1-Gln.

CC -I- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC -I- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO

CC COLLAGEN TYPE XIV (BY SIMILARITY).

CC -I- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR

CC MATRIX (BY SIMILARITY).

CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI;

CC ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-

CC PROCOLLAGEN PEPTIDASE ACTIVITY.

CC -I- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON

CC AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.

CC -I- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT

CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

CC -I- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY

CC SIMILARITY).

CC -I- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF EHLERS-DANLOS

CC SYNDROME TYPE VIIC (EDS-VIIC), A RECESSIVELY INHERITED DISORDER

CC CHARACTERIZED CLINICALLY BY SEVERE SKIN FRAGILITY AND

CC BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN INCOMPLETELY

CC PROCESSED AT THE AMINO TERMINUS.

CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -I- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -I- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.

CC -I- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC -----

EMBL; AJ003125; CAA05880.1; --

DR MEROPS; M12.301; --

DR MIM; 604539; --

DR MIM; 225410; --

DR InterPro; IPR001762; Disintegrin.

DR InterPro; IPR002870; Pep\_M12B\_propep.

DR InterPro; IPR001590; Reprolysin.

DR InterPro; IPR000884; TSP1.

DR InterPro; IPR000130; Zn\_Mtpeptdse.

DR Pfam; PF01562; Pep\_M12B\_propep; 1.

DR Pfam; PF01421; Reprolysin; 1.

DR Pfam; PF00090; TSP1; 4.

DR SMART; SM00209; TSP1; 4.

DR PROSITE; PS50215; ADAM\_MEPRO; 1.

DR PROSITE; PS50092; TSP1; 1.

DR PROSITE; PS00142; ZINC\_PROTEASE; FALSE\_NEG.

DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.

DR KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

DR Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;

DR KW Alternative splicing.

FT SIGNAL 1 29

FT PROPEP 30 253

FT CHAIN 1211

FT ADAMTS-2.

FT METAL 408 408

FT ACT\_SITE 409 409

FT METAL 412 412

FT METAL 418 418

FT DOMAIN 480 560

FT DOMAIN 561 617

FT DOMAIN 618 722

FT SITE 691 693

FT DOMAIN 723 851

FT DOMAIN 852 911

FT DOMAIN 912 974

FT DOMAIN 975 1030

FT DOMAIN 40 43

FT DOMAIN 185 188

FT DOMAIN 112 112

FT CARBOHYD 251 251

FT CARBOHYD 949 949

FT CARBOHYD 993 993

FT CARBOHYD 1031 1031

FT CARBOHYD 1098 1098

FT CARBOHYD 1145 1145

FT CARBOHYD 1150 1150

FT VARSPLIC 544 566

FT VARSPLIC 567 1211

FT MISSING (IN ISOFORM SPNPI).

FT SEQUENCE 1211 AA; 134722 MW; BCEE25C23CAD2D CRC64;

Query Match 5.2%; Score 130; DB 1; Length 1211;

Best Local Similarity 21.8%; Pred. No. 0.021;

Matches 63; Conservative 26; Mismatches 102; Indels 98; Gaps 13;

QY 192 PPRGWDHTAPGHRFTETKD---QPEYDSTDGE-GDMSLWSVSVTCGNGNKRTRSC---244

Db 532 PPLDGTMCAPGKHCFCGHCICLWLPDILKRDGSGWANGSPFGSCSRTCTGTVKFRTRQCDNP 591

QY 245 -----GYACT--ATESRTCDRPNCPGIEDTFRTAATEVSLLAGSEEFNATKLFVDTDC 297

Db 592 HPANGGRTCSGLAYDQLCSRQDCPDLSADFR-----EEQC 627

QY 298 ERWMSCKSEFLKKYMKHVMNDLSPCSCSYTEVAYSTADIFDKRKDFRW-----KDas 352

Db 628 RQW-----DLYFEHGDAAQHHLWLPHEHRDA- 651

QY 353 GPKEKLEIYKPTARYCIRSMLSLESTTLAAHCYCGDNMOLLITRG--KGAGTNLSTEF 410

Db 652 --KERCHLYCESRE--TGEVYSMKRMVHDGTRCSYKDAFSLCVGRGCKRVGCGVIG---704

QY 411 SAEILHYKVDVLPWLICKGWSRYNEARPPNNGOKCTESPSEDIYIKOFQ 459

```
Db 705 SSKQEDKCGV-----CGGDNHCKVVK-----GTFTSPKKHGYIKMFE 743
RESULT 7
TSP2_BOVIN
ID TSP2_BOVIN STANDARD; PRT; 1170 AA.
AC Q95116; Q28180;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 2 precursor (Corticotropin-induced secreted protein)
DE (CISP).
GN THB2 OR TSP2 OR TSP-2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
[1]
SEQUENCE FROM N.A.
RA Danik M., Chinn A., Lafeuillade M., Keramidas M., Aguesse-Germon S.,
RA Penhoat A., Chen H., Mosher D., Chambaz E.M., Feige J.J.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE OF 1-522 FROM N.A.
RX MEDLINE=96331130; PubMed=8698834;
RA Lafeuillade B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
RA Feige J.J.;
RT "Opposite regulation of thrombospondin-1 and corticotropin-induced
RT secreted protein/thrombospondin-2 expression by adrenocorticotrophic
RT hormone in adrenocortical cells";
RL J. Cell. Physiol. 167:164-172(1996).
[3]
SEQUENCE OF 318-831 FROM N.A.
RP TISSUE=ortic endothelium;
RC Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RT "Cloning and sequencing of bovine thrombospondin stimulatory effect of
RT TGF-beta";
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 WVFC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X96540; CAA65385.1; -
DR EMBL; X87620; CAA60952.1; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; WVFC.
DR InterPro; IPR003367; TSP-3.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 9.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSP1; 3.

DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00092; TSP1; 3.
DR PROSITE; PS01208; WVFC; 1.
DR KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 1170 THROMBOSPONDIN 2.
FT DOMAIN 19 232 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 318 375 WVFC.
FT DOMAIN 379 430 TSP TYPE-1 1.
FT DOMAIN 435 491 TSP TYPE-1 2.
FT DOMAIN 492 546 TSP TYPE-1 3.
FT DOMAIN 547 587 EGF-LIKE 1.
FT DOMAIN 588 645 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 646 690 EGF-LIKE 3.
FT DOMAIN 723 758 TSP TYPE-3 1.
FT DOMAIN 759 781 TSP TYPE-3 2.
FT DOMAIN 782 817 TSP TYPE-3 3.
FT DOMAIN 818 840 TSP TYPE-3 4.
FT DOMAIN 841 878 TSP TYPE-3 5.
FT DOMAIN 879 914 TSP TYPE-3 6.
FT DOMAIN 915 950 TSP TYPE-3 7.
FT DOMAIN 951 1170 C-TERMINAL.
FT SITE 926 928 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 266 266 INTERCHAIN (PROBABLE).
FT DISULFID 270 270 INTERCHAIN (PROBABLE).
FT DISULFID 551 562 BY SIMILARITY.
FT DISULFID 556 572 BY SIMILARITY.
FT DISULFID 575 586 BY SIMILARITY.
FT DISULFID 592 608 BY SIMILARITY.
FT DISULFID 599 617 BY SIMILARITY.
FT DISULFID 620 644 BY SIMILARITY.
FT DISULFID 650 663 BY SIMILARITY.
FT DISULFID 657 676 BY SIMILARITY.
FT DISULFID 678 689 BY SIMILARITY.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 582 582 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 708 708 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 936 936 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 535 535 A -> V (IN REF. 3).
FT CONFLICT 748 748 S -> T (IN REF. 3).
SQ SEQUENCE 1170 AA; 129862 MW; 9CF1FBF5B89A051 CRC64;

Query Match 5.1%; Score 128.5; DB 1; Length 1170;
Best Local Similarity 23.1%; Pred. No. 0.027;
Matches 61; Conservative 18; Mismatches 86; Indels 99; Gaps 14;

QY 146 ENKPSWSPVSPDWRWQRSLARANSQDQYXYDSTSDSNFLNP----- 192
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 377 DGEEGHS----PWAETECs---ATCGSGTQGRSCDVTSTNCLGPSIOTRACSLGRCD 429
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 193 -----PRGWDHTAP-----GHRTF-----ETKD-QPEYD 215
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 HRIHQDGGWHSWSPSSCVTCGVGNVTRIRLNCNSPVQPMQGRSKSGSGRETRACQGPC 489
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 216 STDGE-GDWSLWSVCVTCGNGNOKRTRSC-----GYACT--ATESRCDRPNC--- 261
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 490 PVDGRWSPWSPSACTVTCAGGIRIRVCNPSPEPHGGKDCVGGAKEQQMCRKSCPID 549
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 262 -----PGIE-DTERTAAEVSLLAGSEFNATKLFVD-----TDCERWMSCKSEF 307
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 550 GCLSNPCFFGAECSSFPDGSWSCGCPGGFLGNGTHCEDLDECAVVDVC-----F 600
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 308 LKKYMHKVMNDLPS-----CPCSY 326
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 601 ATSKAHRCVNTNPGYHCLPCPPRY 624
RESULT 8
TSPL_BOVIN
ID TSPL_BOVIN STANDARD; PRT; 1170 AA.
AC Q28178; Q28179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 1 precursor.
GN THBS1 OR TSP1 OR TSP-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactylia; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HOLSTEIN; TISSUE=Tooth;
RX MEDLINE=98173773; PubMed=9507054;
RA Ueno A., Yamashita K., Nagata T., Tsurumi C., Miwa Y., Kitamura S.,
RA Inoue H.;
RT "cDNA cloning of bovine thrombospondin 1 and its expression in
RT odontoblasts and predentin.";
RL Biochim. Biophys. Acta 1382:17-22(1998).
RN [2]
RP SEQUENCE OF 1-18 AND 710-1170 FROM N.A.
RC TISSUE=Aortic endothelium;
RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-
CC V/BETA-3 AND ALPHA-1IB/BETA-3. MAY PLAY A ROLE IN DENTINOGENESIS
CC AND/OR MAINTENANCE OF DENTIN AND DENTAL PULP.
CC -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -!- TISSUE SPECIFICITY: ODONTOBLASTS.
CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB005287; BAA21115.1; -.
DR EMBL; X87618; CAA60950.1; -.
DR EMBL; X87619; CAA60951.1; -.
DR HSSP; P35555; IEMO.
DR GlycoSuiteDB; Q28178; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; VWFC.
DR InterPro; IPR003367; tsp_3.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF02412; tsp_3; 16.
DR Pfam; PF00093; vwc; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00209; TSP1; 3.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.

DR DR PROSITE; PS01186; EGF_2; FALSE_NEG.
DR DR PROSITE; PS01208; VWFC; 1.
DR DR PROSITE; PS00992; TSP1; 3.
KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1170
FT DOMAIN 19 232
FT DOMAIN 316 373
FT DOMAIN 379 430
FT DOMAIN 435 491
FT DOMAIN 492 548
FT DOMAIN 549 587
FT DOMAIN 588 645
FT DOMAIN 646 690
FT DOMAIN 723 758
FT DOMAIN 759 781
FT DOMAIN 782 817
FT DOMAIN 818 840
FT DOMAIN 841 878
FT DOMAIN 879 914
FT DOMAIN 915 950
FT DOMAIN 951 1170
FT SITE 926 928
FT DISULFID 270 270
FT DISULFID 274 274
FT DISULFID 551 562
FT DISULFID 556 572
FT DISULFID 575 586
FT DISULFID 592 608
FT DISULFID 599 617
FT DISULFID 620 644
FT DISULFID 650 663
FT DISULFID 657 676
FT DISULFID 678 689
FT DISULFID 248 248
FT CARBOHYD 360 360
FT CARBOHYD 708 708
FT CARBOHYD 1067 1067
FT CARBOHYD 1085 1085
FT CONFLICT 805 805
SQ SEQUENCE 1170 AA; 129533 MW; 0DD6ADF3E5FA031A CRC64;

Query Match 5.1%; Score 127.5; DB 1; Length 1170;
Best Local Similarity 25.6%; Pred. No. 0.032;
Matches 45; Conservative 10; Mismatches 56; Indels 65; Gaps 9;

QY 151 WSPVSPD--WRW--W-----QRLSLARAN-----SGDDYXYDSTSDSNF 189
DB 373 WPSDSADDGWSWSEWTSVTCGNGIQGRSCDILNNRCESVQTRTCHIQECDKRF 432
QY 190 LNPGRGWDHTAPGH-----RTFETK--DQPEYDS 216
DB 433 -KODGWSHSPWSSCSVTCGGDVITRILCNLSPSPQMGKPCGEGRKATKACOKSCPI 491
QY 217 TDGEGDWSLWSVCSVTCGNGNQKRTSC-----GYACT--ATESRTCDRNP 262
DB 492 NGWGWPSPWDICSVTCGGVGQKRSRLCNPKPQFGKDCVGDVTENQICNKQDCP 547

RESULT 9
TSPL_MOUSE
ID TSPL_MOUSE STANDARD; PRT; 1170 AA.
AC P35441;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 1 precursor.
GN THBS1 OR TSP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92128941; PubMed=1774063;  
 RA Lawler J., Duquette M., Ferro P., Copeland N.G., Gilbert D.J.,  
 RA Jenkins N.A.;  
 RT "Characterization of the murine thrombospondin gene.";  
 RL Genomics 11:587-600(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92147683; PubMed=1371115;  
 RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,  
 RA Dixit V.M.;  
 RT "Characterization of mouse thrombospondin 2 sequence and expression  
 during cell growth and development.";  
 RL J. Biol. Chem. 267:3274-3281(1992).  
 RN [3]  
 RP SEQUENCE OF 1-490 FROM N.A.  
 RX MEDLINE=90375546; PubMed=2398070;  
 RA Bornstein P., Alfi D., Devarayalu S., Framson P., Li P.;  
 RT "Characterization of the mouse thrombospondin gene and evaluation of  
 the role of the first intron in human gene expression.";  
 RL J. Biol. Chem. 265:16691-16698(1990).  
 CC -|- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND  
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,  
 LAMININ, TYPE V COLLAGEN AND INTEGRINS ALPHA-V/BETA-1, ALPHA-  
 V/BETA-3 AND ALPHA-1IB/BETA-3.  
 CC -|- SUBUNIT: HOMOPRIMER; DISULFIDE-LINKED.  
 CC -|- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC -|- SIMILARITY: CONTAINS 1 WFDC DOMAIN.  
 CC -|- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.  
 CC -|- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.  
 CC -|- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M62450; AAA50611.1; JOINED.  
 DR EMBL; M62451; AAA50611.1; JOINED.  
 DR EMBL; M62452; AAA50611.1; JOINED.  
 DR EMBL; M62453; AAA50611.1; JOINED.  
 DR EMBL; M62454; AAA50611.1; JOINED.  
 DR EMBL; M62455; AAA50611.1; JOINED.  
 DR EMBL; M62456; AAA50611.1; JOINED.  
 DR EMBL; M62457; AAA50611.1; JOINED.  
 DR EMBL; M62458; AAA50611.1; JOINED.  
 DR EMBL; M62459; AAA50611.1; JOINED.  
 DR EMBL; M62460; AAA50611.1; JOINED.  
 DR EMBL; M62461; AAA50611.1; JOINED.  
 DR EMBL; M62462; AAA50611.1; JOINED.  
 DR EMBL; M62463; AAA50611.1; JOINED.  
 DR EMBL; M62464; AAA50611.1; JOINED.  
 DR EMBL; M62465; AAA50611.1; JOINED.  
 DR EMBL; M62466; AAA50611.1; JOINED.  
 DR EMBL; M62467; AAA50611.1; JOINED.  
 DR EMBL; M62468; AAA50611.1; JOINED.  
 DR EMBL; M62469; AAA50611.1; JOINED.  
 DR EMBL; M62470; AAA50611.1; JOINED.  
 DR EMBL; M62471; AAA50611.1; JOINED.  
 DR PIR; A40558; A40558.  
 DR PIR; B42587; B42587.  
 DR HSSP; P35555; 1EMN.  
 DR MGD; MGI:98737; Thbs1.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001881; EGF\_Ca.

DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR000884; TSP1.  
 DR InterPro; IPR003129; TSPN.  
 DR InterPro; IPR001007; WFDC.  
 DR InterPro; IPR003367; tsp\_3.  
 DR Pfam; PF00008; EGF; 2.  
 DR Pfam; PF02210; TSPN; 1.  
 DR Pfam; PF00090; tsp\_1; 3.  
 DR Pfam; PF02412; tsp\_3; 8.  
 DR Pfam; PF00093; vwc; 1.  
 DR SMART; SM00181; EGF; 3.  
 DR SMART; SM00209; TSP1; 3.  
 DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00214; vwc; 1.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS00092; TSP1; 3.  
 DR PROSITE; PS01208; WFDC; 1.  
 DR Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;  
 KW EGF-like domain; Signal.  
 FT SIGNAL 1 18 POTENTIAL  
 FT CHAIN 19 1170 THROMBOSPONDIN 1  
 FT DOMAIN 19 232 HEPARIN-BINDING (POTENTIAL).  
 FT DOMAIN 316 373 WFDC.  
 FT DOMAIN 379 430 TSP TYPE-1 1.  
 FT DOMAIN 435 491 TSP TYPE-1 2.  
 FT DOMAIN 492 548 TSP TYPE-1 3.  
 FT DOMAIN 549 587 EGF-LIKE 1.  
 FT DOMAIN 588 645 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 646 690 EGF-LIKE 3.  
 FT DOMAIN 723 758 TSP TYPE-3 1.  
 FT DOMAIN 759 781 TSP TYPE-3 2.  
 FT DOMAIN 782 817 TSP TYPE-3 3.  
 FT DOMAIN 818 840 TSP TYPE-3 4.  
 FT DOMAIN 841 878 TSP TYPE-3 5.  
 FT DOMAIN 879 914 TSP TYPE-3 6.  
 FT DOMAIN 915 950 TSP TYPE-3 7.  
 FT DOMAIN 951 1170 C-TERMINAL.  
 FT SITE 926 928 CELL ATTACHMENT SITE (POTENTIAL).  
 FT DISULFID 270 270 INTERCHAIN (PROBABLE).  
 FT DISULFID 274 274 INTERCHAIN (PROBABLE).  
 FT DISULFID 551 562 BY SIMILARITY.  
 FT DISULFID 556 572 BY SIMILARITY.  
 FT DISULFID 575 586 BY SIMILARITY.  
 FT DISULFID 592 608 BY SIMILARITY.  
 FT DISULFID 599 617 BY SIMILARITY.  
 FT DISULFID 620 644 BY SIMILARITY.  
 FT DISULFID 650 663 BY SIMILARITY.  
 FT DISULFID 657 676 BY SIMILARITY.  
 FT DISULFID 678 689 BY SIMILARITY.  
 FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 1025 1025 F -> L (IN REF. 2).  
 SQ SEQUENCE 1170 AA; 129646 MW; 0443E493615E7F06 CRC64;

Query Match

5.1%; Score 127.5; DB 1; Length 1170;

Best Local Similarity 23.3%; Pred. No. 0.032;

Matches 41; Conservative 17; Mismatches 53; Indels 65; Gaps 8;

QY 151 WSVPSPD--WRAWWORSLSLARANSQDQ--YXYDSTSD-----DSNF 189  
 Db 373 WPSDSADDDGHSWSPSEWTSCTCGNGIQGRSCDSLNNRCSSSVOTRCHIQECDKRF 432  
 QY 190 LNPFRGWDHTAPGHRFTET-----KDOPEYDSTDEGD----- 222  
 Db 433 -KQDGGWSHSPWSSCSVTCGDGVITRILCNPSPPOMNGKPCGEARETKACKKDACPI 491  
 QY 223 -----WSLWSVCSVTCGNGNQKRTSC-----GYACT--ATESRTCDRNCP 262  
 Db 492 NGGWPSPWDICSVTCGGGVRRSLCNNTPTQFGGKDCVGDVTENQVCNKQDCP 547



```

RESULT 10
TSPL_HUMAN
ID TSPL_HUMAN STANDARD; PRT; 1170 AA.
AC P07996;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 1 precursor.
GN THBS1 OR TSPL OR TSP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Endothelial cells;
RA Lawler J., Hynes R.O.;
RX MEDLINE=87057617; PubMed=2430973;
RT "The structure of human thrombospondin, an adhesive glycoprotein with
RT multiple calcium-binding sites and homologies with several different
RT proteins.";
RL J. Cell Biol. 103:1635-1648(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89139590; PubMed=2918029;
RA Hennessy S.W., Frazier W.A., Kim D.D., Deckwerth T.L.,
RA Baumgartel D.M., Rotwein P., Frazier W.A.;
RT "Complete thrombospondin mRNA sequence includes potential regulatory
RT sites in the 3' untranslated region.";
RL J. Cell Biol. 108:729-736(1989).
RN [3]
RP SEQUENCE OF 1-397 FROM N.A.
RX MEDLINE=87157592; PubMed=3030396;
RA Kobayashi S., Eden-McCuchan F., Framson P., Bornstein P.;
RT "Partial amino acid sequence of human thrombospondin as determined by
RT analysis of cDNA clones: homology to malarial circumsporozoite
RT proteins.";
RL Biochemistry 25:8418-8425(1986).
RN [4]
RP SEQUENCE OF 1-374 FROM N.A.
RX MEDLINE=86287276; PubMed=3461443;
RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
RT "Characterization of a cDNA encoding the heparin and collagen binding
RT domains of human thrombospondin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
RN [5]
RP SEQUENCE OF 1-166 FROM N.A.
RX MEDLINE=89291870; PubMed=2544587;
RA Laherty C.D., Gierman T.M., Dixit V.M.;
RT "Characterization of the promoter region of the human thrombospondin
RT gene. DNA sequences within the first intron increase transcription.";
RL J. Biol. Chem. 264:11222-11227(1989).
RN [6]
RP SEQUENCE OF 1028-1170 FROM N.A.
RA la Fleur M., Jobin C., Gauthier J., Kreis C.G.;
RT Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
RN [7]
RP CARBOHYDRATE-LINKAGE SITES W-385; S-394; W-438; W-441; T-450; W-498
AND T-507.
RC TISSUE-Platelet;
RX PubMed=11067851;
RA Hofsteenge J., Huwiler K.G., Macek B., Hess D., Lawler J.,
RA Mosher D.F., Peter-Katalinic J.;
RT "C-mannosylation and O-fucosylation of the thrombospondin type 1
RT module.";
RL J. Biol. Chem. 276:6485-6498(2001).
CC -!- FUNCTION: Adhesive glycoprotein that mediates cell-to-cell and
CC cell-to-matrix interactions. Can bind to fibrinogen, fibronectin,
CC laminin, type V collagen and integrins alpha-V/beta-1, alpha-
CC V/beta-3 and alpha-IIIb/beta-3.
CC -!- SUBUNIT: Homotrimer; disulfide-linked.

```

---

```

CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -!- SIMILARITY: CONTAINS 3 WFEC DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M25631; AAA36741.1; -
CC EMBL; X04665; CAA28370.1; -
CC EMBL; X14787; CAA32889.1; -
CC EMBL; J04835; AA61178.1; -
CC EMBL; M99425; AAB59366.1; -
CC PIR; A05172; A05172.
CC PIR; A25812; A25812.
CC PIR; A26155; A26155.
CC PIR; A30140; A30140.
CC PIR; A34274; A34274.
CC HSP; P35555; IEMN.
CC GlycoSuiteDB; P07996; -
CC MIM; 188060; -
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001881; EGF-Ca.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR003129; TSPN.
CC InterPro; IPR001007; WFEC.
CC InterPro; IPR003367; tsp_3.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF02210; TSPN; 1.
CC Pfam; PF00090; tsp_1; 3.
CC Pfam; PF02412; tsp_3; 8.
CC Pfam; PF00093; vwc; 1.
CC SMART; SM00181; EGF; 3.
CC SMART; SM00209; TSP1; 3.
CC SMART; SM00214; TSPN; 1.
CC SMART; SM00214; vwc; 1.
CC PROSITE; PS00022; EGF_1; FALSE_NEG.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS50092; TSP1; 3.
CC PROSITE; PS01208; WFEC; 1.
CC Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW EGF-like domain; Signal.
FT SIGNAL 1 18
FT CHAIN 19 1170
FT DOMAIN 19 232.
FT DOMAIN 316 373
FT DOMAIN 379 430
FT DOMAIN 435 491
FT DOMAIN 492 548
FT DOMAIN 549 587
FT DOMAIN 588 645
FT DOMAIN 646 690
FT DOMAIN 723 758
FT DOMAIN 759 781
FT DOMAIN 782 817
FT DOMAIN 818 840
FT DOMAIN 841 878
FT DOMAIN 879 914
FT DOMAIN 915 950
FT DOMAIN 951 1170
FT SITE 926 928
FT DISULFID 270 270
FT DISULFID 274 274
FT DISULFID 551 562
FT DISULFID 556 572
FT DISULFID 575 586

```



```

FT DISULFID 592 608 BY SIMILARITY.
FT FT DISULFID 599 617 BY SIMILARITY.
FT FT DISULFID 620 644 BY SIMILARITY.
FT FT DISULFID 650 663 BY SIMILARITY.
FT FT DISULFID 657 676 BY SIMILARITY.
FT FT DISULFID 678 689 BY SIMILARITY.
FT FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .).
FT FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT CARBOHYD 385 385 C-LINKED (MAN).
FT FT CARBOHYD 394 394 /FTid=CAR_000205.
FT FT CARBOHYD 438 438 O-LINKED (FUC. . .).
FT FT CARBOHYD 441 441 /FTid=CAR_000206.
FT FT CARBOHYD 441 441 C-LINKED (MAN).
FT FT CARBOHYD 450 450 /FTid=CAR_000207.
FT FT CARBOHYD 498 498 C-LINKED (MAN).
FT FT CARBOHYD 507 507 /FTid=CAR_000208.
FT FT CARBOHYD 708 708 O-LINKED (FUC. . .).
FT FT CARBOHYD 1067 1067 /FTid=CAR_000209.
FT FT CONFLICT 84 84 C-LINKED (MAN).
FT FT CONFLICT 523 523 /FTid=CAR_000210.
FT FT CONFLICT 1170 AA; 129412 MW; 69B3EDE5AE3A395E CRC64;
SQ SEQUENCE 1170 AA; 129412 MW; 69B3EDE5AE3A395E CRC64;

Query Match 5.0%; Score 127; DB 1; Length 1170;
Best Local Similarity 44.4%; Pred. No. 0.035;
Matches 24; Conservative 7; Mismatches 13; Indels 10; Gaps 2;

Qy 219 GEGDWSLWCVGCVGNGNQRTRSC-----GYACT--ATESRTCDRNCP 262
Db 494 GGWSPMDICSVTCGGVQRRSLCNPNPTQFGKDCVGDVTENQCNKQDCP 547

RESULT 11
ATS2_BOVIN STANDARD; PRT; 1205 AA.
AC P79331;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI).
GN ADAMTS2 OR NPI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=97225960; PubMed=9122202;
RA Colige A., Li S.W., Sieron A.L., Nusgens B.V., Prockop D.J.,
RA Lapiere C.M.;
RT "cDNA cloning and expression of bovine procollagen I N-proteinase: a
RT new member of the superfamily of zinc-metalloproteinases with binding
RT sites for cells and other matrix components.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).
RN [2]
RN PARTIAL SEQUENCE.
RX MEDLINE=95348096; PubMed=7622483;
RA Colige A., Beschin A., Sanyon B., Goebels Y., Van Beeumen J.,
RA Nusgens B.V., Lapiere C.M.;
RT "Characterization and partial amino acid sequencing of a 107-kDa
RT procollagen I N-proteinase purified by affinity chromatography on
RT immobilized type XIV collagen.";

```

```

J. Biol. Chem. 270:16724-16730(1995).
-|- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
COLLAGEN BIOSYNTHESIS.
-|- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
alpha-1(I) at Pro-I-Gln and of alpha-1(II) and alpha-2(I) chains
at Ala-I-Gln.
-|- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
-|- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
COLLAGEN TYPE XIV.
-|- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
MATRIX (BY SIMILARITY).
-|- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVELS
IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS
AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS
WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.
-|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
-|- PTM: THE N-TERMINUS IS BLOCKED.
-|- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
SIMILARITY).
-|- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF DERMATOSPARAXIS, A
RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN
FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN
INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.
-|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
-|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
-|- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X96389; CAA65253.1; -
InterPro; IPR001762; Disintegrin.
InterPro; IPR002870; Pep_M12B_propep.
InterPro; IPR001590; Reprolysin.
InterPro; IPR000884; TSPI.
InterPro; IPR000130; Zn_MTpeptidse.
Pfam; PF01562; Pep_M12B_propep; 1.
Pfam; PF01421; Reprolysin; 1.
SMART; SM00090; tsp_1; 4.
SMART; SM00209; TSPI; 4.
PROSITE; PS50215; ADAM_MEPRO; 1.
PROSITE; PS50092; TSPI; 1.
PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 28 POTENTIAL.
FT PROPEP 29 253 BY SIMILARITY.
FT CHAIN 254 1205 ADAMTS-2.
FT METAL 402 402 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 403 403 BY SIMILARITY.
FT METAL 406 406 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 474 554 DISINTEGRIN-LIKE.
FT DOMAIN 555 611 TSP TYPE-1 1.
FT DOMAIN 612 716 CYS-RICH.
FT DOMAIN 717 845 SPACER.
FT DOMAIN 846 905 TSP TYPE-1 2.
FT DOMAIN 906 968 TSP TYPE-1 3.
FT DOMAIN 969 1024 TSP TYPE-1 4.
FT SITE 685 687 CELL ATTACHMENT SITE (POTENTIAL).
FT DOMAIN 31 35 POLY-ALA.
FT DOMAIN 177 180 POLY-GLU.
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 943 943 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 987 987 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1025 1025 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1092 1092 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1144 1144 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1205 AA; 133887 MW; 7B5B232A45320371 CRC64;

Query Match 4.9%; Score 124; DB 1; Length 1205;  
 Best Local Similarity 20.9%; Pred. No. 0.063;  
 Matches 64; Conservative 26; Mismatches 84; Indels 132; Gaps 15;

QY 192 PPRGWDHTAPGHRTE-----TKDPEYDSTDGEGDWSLWS---VCSVTCNGNGNKRTR 242  
 DB 526 PPLDGTWCAPGKHCFKGHCITWLTPDILKRD-----GNWGAWSFGSCSTCGTGVKFRTR 580  
 QY 243 SC-----GYACT--ATESRTCDRNCPCGIEDTFTTAATEVSLLAGSEEFNATKLFEV 292  
 DB 581 QCDNPHPANGRTCSGLAYDQLCNSQDCPDALADR----- 617  
 QY 293 DTDSCERWMSCKSEFLKXKMKVNDLPSCPSYPTVEVAYSTADIFRIKRDPRW----- 348  
 DB 618 -EEOCRQW-----DLYPEHGDAQHHWLPHE 641  
 QY 349 -KDAAGPKKLEIV---KPTARYCIRMSLSLESTTLAAQHCVCYGDNNQLITRG--KGAGT 402  
 DB 642 HRDA---KERCHLYCESKETG-----EVSMKRMVHDGTRCSYKDAFSLVCGDCRVKVC 693  
 QY 403 PNLISTEFSAE-----LHYKVDVLPWLICKGDMWSRYNEARPPNNGOKCTESPDES 453  
 DB 694 DGVIGSSKQEDKCGVCGDGNHCKV-----YKGTFSR-----SPKGLG 731  
 QY 454 YIKOFQ 459  
 DB 732 YIKMFE 737

## RESULT 12

## ATSA\_HUMAN

ID ATSA4\_HUMAN STANDARD; PRT; 837 AA.  
 AC 075173; Q9UN83;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
 DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)  
 DE (ADMP-1).  
 GN ADAMTS4 OR KIAA0688.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain.  
 RX MEDLINE=98403880; PubMed=9734811;  
 RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,  
 RA Kotani H., Nomura N., Ohara O.;  
 RA "Prediction of the coding sequences of unidentified human genes. X.  
 RA The complete sequences of 100 new cDNA clones from brain which can  
 RA code for large proteins in vitro.";  
 RL DNA Res. 5:169-176(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99286303; PubMed=10356395;  
 RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,  
 RA Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,  
 RA Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,  
 RA Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,  
 RA Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,

RA Trzaskos J.M., Arner E.C.;  
 RT "Purification and cloning of aggrecanase-1: a member of the ADAMTS  
 RT family of proteins.";  
 RL Science 284:1664-1666(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Sawaji Y., Nagase H., Saklatvala J., Clark A.R.;  
 RT "ADAMTS-4 genomic locus.";  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.  
 RX MEDLINE=20400518; PubMed=10827174;  
 RA Tortorella M.D., Pratta M.A., Liu R.-Q., Abbaszade I., Ross H.,  
 RA Burn T.C., Arner E.C.;  
 RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for  
 RT aggrecan substrate recognition and cleavage.";  
 RL J. Biol. Chem. 275:25791-25797(2000).  
 CC -!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE  
 CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE  
 CC DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE  
 CC CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN  
 CC ALZHEIMER'S DISEASE.  
 CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393  
 CC site.  
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR  
 CC MATRIX (BY SIMILARITY).  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED  
 CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.  
 CC -!- INDUCTION: BY INTERLEUKIN-1.  
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT  
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.  
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
 CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO AS ADAMTS2.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB014588; BAA31663.1; -;  
 DR EMBL; AF148213; AAD41494.1; -;  
 DR EMBL; AY044847; AAL02262.1; -;  
 DR HSSP; P34179; ILAG.  
 DR MEROPS; M12.221; -;  
 DR TIM; 603876; -;  
 DR InterPro; IPR001762; Disintegrin.  
 DR InterPro; IPR001590; Reprolysin.  
 DR InterPro; IPR000884; TSP1.  
 DR InterPro; IPR000130; Zn\_MTpeptase.  
 DR Pfam; PF01421; Reprolysin; 1.  
 DR Pfam; PF00090; tsp\_1; 1.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR PROSITE; PS50215; ADAM\_MEPRO; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 DR HydroLase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Extracellular matrix.  
 FT SIGNAL 1 51 POTENTIAL.  
 FT PROPEP 52 212  
 FT CHAIN 213 837  
 FT SITE 194 194  
 FT METAL 361 361  
 FT ACT\_SITE 362 362  
 FT METAL 365 365  
 FT METAL 371 371

```

FT DOMAIN 437 519 DISINTEGRIN-LIKE.
FT DOMAIN 520 576 TSP TYPE-1.
FT DOMAIN 577 685 CYS-RICH.
FT DOMAIN 686 837 SPACER.
FT DOMAIN 247 252 POLY-ALA.
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 77 77 A -> T (IN REF. 1).
FT CONFLICT 636 636 R -> Q (IN REF. 3).
FT CONFLICT 682 682 G -> R (IN REF. 3).
SQ SEQUENCE 837 AA: 90224 MW: 5DF9C9AC137DF41F CRC64;

Query Match 4.8%; Score 122; DB 1; Length 837;
Best Local Similarity 22.3%; Pred. No. 0.057;
Matches 61; Conservative 21; Mismatches 88; Indels 104; Gaps 11;

QY 104 LLDLPNFPDLSRADINGQPNIOVTIEVDGPDSEADKDOH-----PENKPSWSVPSPDWR 159
Db 424 LLDKPEAP-----LHLPVTF---PGKDYADADQCQUTFGDPSRHCPCQLPPPCAA 469
QY 160 ANWQSLSLARANSQDQXYDSTSDSNFLNPPRGWDHTAPGHRTFFTKDOPEYDST-- 217
Db 470 LMC-----SCHLN-----CHAMCQTKHSPWADCTPC 495
QY 218 -----DGEDWSLWSVCTCGNGNQKTRSC----- 244
Db 496 GPAQACMGRCRLHMDQLQDNIPQAGGPGWPGWGCSTCGGGVQFSSRDCTRPVPRNG 555
QY 245 GYACTA--TESTCDRPNCP-GIEDTFRYAATEVSLLAGSEEFNATKLEFVDTDSCRW- 300
Db 556 GRYCEGRRTRFRSCNTEDCPTGSALTFR---EEQCAAYNHRTDLFKSPGPMQDWPVRYT 611
QY 301 -----MSCSKSEFLKMYMKVNDLPSCPCS 325
Db 612 GVAPQDCKLTQOARLGYVYVLEPRVVDGTGTPCS 645

RESULT 13
TSP2_MOUSE
ID TSP2_MOUSE STANDARD; PRT: 1172 AA.
AC Q03350;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 2 precursor.
GN THBS2 OR TSP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147683; PubMed=1371115;
RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
RA Dixit V.M.;
RT "Characterization of mouse thrombospondin 2 sequence and expression
RT during cell growth and development.";
RL J. Biol. Chem. 267:3274-3281(1992).
RN [2]
RP SEQUENCE OF 1-873 FROM N.A.
RX MEDLINE=91302287; PubMed=1712771;
RA Bornstein P., O'Rourke K., Wikstrom K., Wolf F.W., Katz R., Li P.,
RA Dixit V.M.;
RT "A second, expressed thrombospondin gene (Thbs2) exists in the mouse
RT genome.";
RL J. Biol. Chem. 266:12821-12824(1991).
CC -|- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -|- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -|- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -|- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -|- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.

```

```

CC -|- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -|- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L07803; AAA53064.1; -
CC EMBL; M64866; AAA40432.1; -
CC PIR; A42587; A42587.
CC PIR; A39851; A39851.
CC HSP; P35555; 1EMN.
CC MGD; MGI:98738; Thbs2.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR001791; Laminin_G.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR003129; TSPN.
CC InterPro; IPR001007; VWFC.
CC InterPro; IPR003367; tsp_3.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF02210; TSPN; 1.
CC Pfam; PF00090; tsp_1; 3.
CC Pfam; PF02412; tsp_3; 9.
CC Pfam; PF00093; vwc; 1.
CC SMART; SM00181; EGF; 3.
CC SMART; SM00209; TSP1; 3.
CC SMART; SM00210; TSPN; 1.
CC SMART; SM00214; VWC; 1.
CC PROSITE; PS00022; EGF_1; FALSE_NEG.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS50092; TSP1; 3.
CC PROSITE; PS01208; VWFC; 1.
CC KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
FT SIGNAL 1 18
FT CHAIN 19 1172
FT DOMAIN 19 232 THROMBOSPONDIN 2.
FT DOMAIN 318 375 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 381 432 VWFC.
FT DOMAIN 437 493 TSP TYPE-1 1.
FT DOMAIN 494 548 TSP TYPE-1 2.
FT DOMAIN 549 589 TSP TYPE-1 3.
FT DOMAIN 590 647 EGF-LIKE 1.
FT DOMAIN 648 692 EGF-LIKE 2.
FT DOMAIN 725 760 EGF-LIKE 3.
FT DOMAIN 761 783 TSP TYPE-3 1.
FT DOMAIN 784 819 TSP TYPE-3 2.
FT DOMAIN 820 842 TSP TYPE-3 3.
FT DOMAIN 843 880 TSP TYPE-3 4.
FT DOMAIN 881 916 TSP TYPE-3 5.
FT DOMAIN 917 952 TSP TYPE-3 6.
FT DOMAIN 953 1172 TSP TYPE-3 7.
FT SITE CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 266 266 INTERCHAIN (PROBABLE).
FT DISULFID 270 270 INTERCHAIN (PROBABLE).
FT DISULFID 553 564 BY SIMILARITY.
FT DISULFID 558 574 BY SIMILARITY.
FT DISULFID 577 588 BY SIMILARITY.
FT DISULFID 594 610 BY SIMILARITY.
FT DISULFID 601 619 BY SIMILARITY.
FT DISULFID 622 646 BY SIMILARITY.
FT DISULFID 652 665 BY SIMILARITY.
FT DISULFID 659 678 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

[illegible]

Db 385 DSEEGSPWSDWTKCSVTCSGTOMRGRSCDVTRSACTPHIQTRMCSFKKC 436

RESULT 15

AT54\_RAT

ID AT54\_RAT STANDARD; PRT; 630 AA.

AC Q9ESP7; Q9ESP8; Q9ESP6;

DT 01-MAR-2002 (Rel. 41, Created)

DT 01-MAR-2002 (Rel. 41, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1) (Fragment).

DE ADAMTS4.

GN Rattus norvegicus (Rat).

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Wistar; TISSUE=Brain;

RX MEDLINE=20415831; PubMed=10961658;

RA Satoh K.; Suzuki N.; Yokota H.;

RT "ADAMTS-4 (a disintegrin and metalloproteinase with thrombospondin motifs) is transcriptionally induced in beta-amyloid treated rat astrocytes.";

RL Neurosci. Lett. 289:177-180(2000).

CC -1- FUNCTION: CLEAVES AGGECAN. A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES.

CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393 site.

CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: ONLY EXPRESSED IN BRAIN.

CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).

DR EMBL: AB042272; BAB16474.1; -

DR EMBL: AB042271; BAB16473.1; -

DR EMBL: AB042273; BAB16475.1; -

DR InterPro: IPR001590; Reprolysin.

DR InterPro: IPR000884; TSP1.

DR Pfam: PF00090; tsp.1; 1.

DR SMART: SM00209; TSP1. 1.

DR PROSITE: PS50215; ADAM\_MEPRO; 1.

DR PROSITE: PS50092; TSP1. 1.

DR PROSITE: PS00142; ZINC\_PROTEASE; 1.

DR PROSITE: PS00427; DISINTEGRIN.1; FALSE NEG.

KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;

Extracellular matrix.

FT NON\_TER 1 1

FT PROPEP <1 5 BY SIMILARITY.

FT CHAIN 6 630 ADAMTS-4.

FT METAL 154 154 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT\_SITE 155 155 BY SIMILARITY.

FT METAL 158 158 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 164 164 ZINC (CATALYTIC) (BY SIMILARITY).

FT DOMAIN 233 303 DISINTEGRIN-LIKE.

FT DOMAIN 316 367 TSP TYPE-1.

FT DOMAIN 368 478 CYS-RICH.

FT DOMAIN 479 630 SPACER.

FT DOMAIN 40 45 POLY-ALA.

FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 630 AA; 68384 MW; 63A428753167C7EF CRC64;

Query Match 4.7%; Score 118; DB 1; Length 630;

Best Local Similarity 21.9%; Pred. No. 0.082;

Matches 60; Conservative 19; Mismatches 91; Indels 104; Gaps 10;

QY 104 LLDLPNFPDLKADINGQXPNQIQTIEVVDGPDSEADKQDQH-----PENKPSWSVPSPDWR 159

Db 217 LLDKPEAP-----LHLPVTF---PGKDYDADROCOLTEGPDSSHCPLPPCAA 262

QY 160 AWMORSILARANSQDQDYXYDSTSDSDSNFLNPPRGWDHTAPGHRTETKDKQDEYDST-- 217

Db 263 LMCFGHLN-----GHAMCQTKHSPWADGTPC 288

QY 218 -----DGEQDWSLWSVCSVTCGNGNOKRTRSC----- 244

Db 289 GPAQACMGGRCLHVDQLKDFNIPQAGGWPMPGMDCSRCTCGGVQVSSRDCTKPVPRNG 348

QY 245 GYACTA--TESRTCDRPNCP-GIEDTFTATAATEVSLLAGSEEFNATKLFVEYDTSCEERW- 300

Db 349 GKYCEGRRTTPRSCTNKNCPHGSALTER-----EEQCAAYNHRTDLFKSFPGPMDWVPRYT 404

QY 301 -----MSCKSEFLKKYMHKVMNDLPSCPCS 325

Db 405 GVAPRDQCKLTQARALGYVYVLEPRVADGTPCS 438

Search completed: April 25, 2002, 17:27:30

Job time: 9602 sec

